Fehlerabschätzungen bei bibliometrischen Analysen

Masterarbeit

zur Erlangung des akademischen Grades Master of Library and Information Science

M.A. LIS

vorgelegt der Technischen Hochschule Köln

Fakultät für Informations- und Kommunikationswissenschaften

am 9. Oktober 2018

von Dr. Felix Schmidt

Gutachter:

1. Dr. Dirk Tunger, Forschungszentrum Jülich, Jülich

2. Prof. Dr. Simone Fühles-Ubach, Technische Hochschule Köln, Köln

Zusammenfassung

Die vorliegende Masterarbeit beschäftigt sich mit Fehleranalysen bei bibliometrischen Kennzahlen. Dabei werden die Auswirkungen von Fehlern in der bei der Analyse verwendeten Datengrundlage auf die Ergebnisse untersucht. Dies wird am Beispiel einiger Kennzahlen verschiedenen Typs durchgeführt. Ziel dabei ist es, eine Aussage darüber treffen zu können, inwieweit Ergebnisse aus bibliometrischen Analysen durch Fehler in den zugrundeliegenden Zitationsdatenbanken verfälscht werden können.

Als Datengrundlage werden Auszüge aus der Datenbank des Kompetenzzentrums Bibliometrie verwendet. In diese werden statistisch generierte Fehlerkonfigurationen unterschiedlicher Fehlerwahrscheinlichkeiten implementiert und bestimmt, wie stark die daraus resultierenden Kennzahlen schwanken. Eine statistische Analyse der auftretenden Verteilungen erlaubt es dann, Aussagen über die Stabilität der Ergebnisse zu treffen. Dabei werden zwei verschiedene Arten von Fehlern näher untersucht.

Analysiert wird zunächst der h-Index, eine meist für Personen verwendete Kennzahl, die sehr verbreitet ist. Anschließend wird die entwickelte Methodik auf normalisierte Indikatoren angewendet. Während bei der Untersuchung des h-Index jedoch künstlich generierte Publikationssets fiktiver Autoren verwendet werden, werden die Untersuchungen zu den normalisierten Indikatoren am realen Beispiel der Universität Duisburg-Essen durchgeführt.

Insgesamt soll die in dieser Arbeit entwickelte und vorgestellte Methodik einen Anstoß dazu liefern, dass die Genauigkeit und Aussagekraft der Ergebnisse bibliometrischer Indikatoren genauer hinterfragt und untersucht wird.

Schlagwörter: Bibliometrie, Zitationsdatenbanken, Fehleranalyse, h-Index, Normalisierte Indikatoren

Abstract

This Master's thesis deals with error analyses in bibliometric indicators. The implications of errors in the used database on the results are investigated by taking the example of various bibliometric indicators of different kinds. The goal is to allow for propositions on how far the results of biblimetric analyses can be distorted by errors in the underlying citation data.

Data from the bibliometric database of the Competence Centre for Bibliometrics are used as a toy model. Different kinds of statistical errors with various error probabilities are implemented and investigated on how the indicators fluctuate due to this. A statistical analysis of the resulting distributions allows for statements on the stability of the results.

The first analysed case is the h-index, an indicator frequently used for individual scientists. Subsequently, the developed method is applied to normalised indicators. Whereas for the h-index artificially generated publication profiles of imaginary scientists are generated and used, the investigations of the normalised indicators are performed using the realistic example of the Univerity of Duisburg-Essen.

Altogether the method developed and presented in this thesis shall contribute to investigations of the acurracy and validity of bibliometric indicators in greater detail.

Tags: bibliometrics, citation databases, error analysis, h-Index, normalised indicators

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

1. Einleitung

Bibliometrische Analysen werden von immer größerer Bedeutung für die moderne Forschungslandschaft.[1] So werden sie zum Beispiel bei Berufungsverfahren, der Drittmittelvergabe und zum Benchmarking verschiedener Forschungsinstitutionen eingesetzt. Dabei wird eine Vielzahl von Kennzahlen verwendet, welche die Rezeption der Publikationen von einzelnen Wissenschaftlern[2], Institutionen, Ländern, Verlagen oder Zeitschriften beschreiben soll. Prominenteste Beispiele dafür sind der h-Index, der *Journal Impact Factor* sowie die meist aus mehreren Indikatoren abgeleiteten Positionen in Rankings. Es gibt jedoch noch viele weitere weniger bekannte Indikatoren, deren Berechnung teilweise sehr komplex ist, bei denen zum Beispiel eine Normalisierung implementiert wird, um eine Vergleichbarkeit über verschiedene Forschungsgebiete zu erzielen. Beispiele dafür sind der *Field Weighted Citation Impact* sowie der J-Faktor.

Die Berechnungen dieser bibliometrischen Kennzahlen erfolgen anhand gegebener Datenbasen, die Publikationseinträge verzeichnen und diese untereinander verlinken. Besonders häufig handelt es sich um eine der beiden kommerziellen Zitationsdatenbanken *Web of Science* oder *Scopus*. Aber auch andere Datengrundlagen, wie zum Beispiel institutionelle Bibliographien oder spezielle (disziplinspezifische) Fachdatenbanken werden zur Berechnung von Metriken verwendet. Es ist davon auszugehen, dass alle verwendeten Datengrundlagen fehlerhafte Einträge in verschiedener Art und Häufigkeit enthalten.

[1] Hicks u. a., „Bibliometrics: The Leiden Manifesto for research metrics“.

[2] Wenn in dieser Arbeit ausschließlich die männliche oder weibliche Form verwendet wird, so dient dies ausschließlich der Lesbarkeit und Einfachheit. Es seien stets Personen des jeweils anderen Geschlechts miteinbezogen, sofern nicht ausdrücklich anders erwähnt.

Häufige Fehler sind dabei Publikationseinträge, die

- einer falschen Institution zugeordnet sind,
- einem falschen Autor zugeordnet sind,
- gänzlich fehlen oder
- fehlerhaft mit anderen Einträgen verlinkt sind (beispielsweise durch fehlende Referenzen).

Insbesondere letzteres hat deutlichen Einfluss auf die Zitationszahlen und somit auf die damit verbundenen bibliometrischen Kennzahlen, die oft zur Evaluation und zu Rankings herangezogen werden (h-Index, Impact-Faktor, J-Faktor). Es gibt eine Reihe von Untersuchungen dazu, wie häufig diese Fehler in den verschiedenen Datenbanken auftreten.[3] Wenig wurden bislang jedoch die Auswirkungen dieser Fehler auf die darauf aufbauende Berechnung von bibliometrischen Kennzahlen und Rankings untersucht. In der Praxis werden bibliometrische Kennzahlen meist ohne Abschätzung der oben genannten Fehler angegeben, in einigen Fällen sogar auf mehrere Nachkommastellen genau, wie zum Beispiel bei Impact-Faktoren mit vier signifikanten Stellen.[4] Diese Genauigkeit kann rechnerisch aufgrund der beschriebenen Ungenauigkeiten der Datengrundlagen nicht nachgewiesen werden und es ist davon auszugehen, dass sie nicht immer eingehalten werden kann.
Diese Arbeit soll daher dazu beitragen, dass künftig vermehrt Abschätzungen für die Genauigkeit der berechnete Kennzahlen in bibliometrischen Analysen angegeben werden, wie es in anderen Bereichen der Wissenschaft bei der Durchführung von Messungen (das schließe Experimente und Simulationen ein) bereits üblich ist.[5]

[3] Siehe zum Beispiel Olensky, „Data accuracy in bibliometric data sources and its impact on citation matching“, und die darin angegebene Literatur.
[4] Minnick, *A closer look at the Journal Impact Factor numerator*.
[5] DIN 1319-1 (1995-01-00), *Grundlagen der Meßtechnik - Teil 1: Grundbegriffe*.

Bibliometrische Fehleranalysen laufen meist in zwei Schritten ab: Zunächst müssen die Fehlerhäufigkeiten der jeweiligen Datengrundlage bestimmt werden. Dies wird meist händisch durch Auswahl und Überprüfung geeigneter Stichproben und anschließende Extrapolation der Ergebnisse auf die Gesamtheit durchgeführt. Alternativ kann dies auch automatisiert durch einen Abgleich verschiedener Datenbanken gegeneinander, beziehungsweise durch Vergleich der Datenbankeinträge mit pdf-Dateien oder Webseiten der Verlage geschehen. Im zweiten Schritt wird darauf aufbauend untersucht, welche Auswirkungen diese Fehler auf die Berechnung verschiedener Metriken haben. In der vorliegenden Ausarbeitung soll dieser zweite Schritt näher untersucht werden.

Der Rest dieser Arbeit gliedert sich folgendermaßen: In den folgenden Unterkapiteln wird eine kurze Einführung in das Thema Bibliometrie gegeben. Dabei sollen nur die für diese Arbeit relevanten Begriffe kurz vorgestellt werden. Für alles Weiterführende sei auf die zahlreich vorhandene Literatur verwiesen.[6] [7] [8] Insbesondere werden die in dieser Arbeit untersuchten bibliometrischen Kennzahlen vorgestellt, auf die Datengrundlage eingegangen und eine Übersicht über die Literatur zum Thema Fehleranalyse bei bibliometrischen Indikatoren gegeben.
Im Hauptteil folgt dann die Vorstellung der grundlegenden Vorgehensweise sowie die eigentlichen Untersuchungen und Ergebnisse zu den Indikatoren – zunächst zum h-Index, dann zu den zwei analysierten normalisierten Metriken. Anschließend folgt eine Diskussion der wesentlichen Resultate sowie ein Ausblick auf mögliche weiterführende Untersuchungen samt Erweiterungs- und Vertiefungsmöglichkeiten der hier durchgeführten Analysen.
Im Anhang sind zudem die wesentlichen Teile des Quellcodes der verwendeten Programme abgebildet und kurz erläutert.

[6] Haustein und Tunger, „Sziento- und bibliometrische Verfahren“.
[7] Gimpl, „Evaluation von ausgewählten Altmetrics-Diensten für den Einsatz an wissenschaftlichen Bibliotheken“.
[8] Havemann, *Einführung in die Bibliometrie*.

1.1. Bibliometrie

Das Themenfeld Bibliometrie hat in der zweiten Hälfte des 20. Jahrhunderts als Methodik für die Bewertung der Quantität der wissenschaftlichen Produktion zunehmend an Bedeutung gewonnen. Ursächlich dafür waren Reaktionen auf das exponentielle Wachstum wissenschaftlicher Publikationen von Seiten der Forschungsförderer und der Wissenschaftler: zum einen kam der Wunsch nach belastbaren Bewertungskriterien und aussagekräftigen Evaluationsergebnissen auf, zum anderen reagierten Wissenschaftler darauf, indem sie ihre Arbeitsweise an diese Entwicklung immer weiter anpassten. Zudem erhofften die Forscher sich durch bibliometrische Methoden Orientierung in der zunehmenden Informationsflut.[9] Im Gegensatz zur ebenfalls gebräuchlichen Evaluation durch Peer-Reviews nutzt die Bibliometrie mathematisch-statistische Evaluationsmethoden, um die wissenschaftliche Produktivität und Produktion in Form von Publikationen sowie deren Rezeption in Form von Zitationen zu analysieren und so mittels möglichst objektiver Kriterien Aussagen dazu treffen zu können.[10] Neben der Auswertung der Anzahl der wissenschaftlichen Publikationen (zum Beispiel Monographien, Aufsatzsammlungen, Zeitschriften, elektronische Medien) werden auch die Zitationen, Kozitationen und Kopublikationen ausgewertet, um quantitative Aussagen über einzelne Wissenschaftler, Institute, Einrichtungen wie Universitäten oder ganze Länder zu treffen, die beispielsweise mittels Ranglisten oder Diagrammen dargestellt werden können.[11] [12]
Auf Seiten der Wissenschaft führt das häufig zu einer gewissen Skepsis gegenüber der Bibliometrie, da befürchtet wird, aufgrund fehlerhafter Methoden benachteiligt zu werden:

> Nowadays many scientists feel uneasy due to a growing pressure to quantify

[9] Jokić und Ball, *Qualität und Quantität wissenschaftlicher Veröffentlichungen : bibliometrische Aspekte der Wissenschaftskommunikation*, S. 7–8.
[10] Jokić und Ball, *Qualität und Quantität wissenschaftlicher Veröffentlichungen : bibliometrische Aspekte der Wissenschaftskommunikation*, S. 10.
[11] Jovanović, „Eine kleine Frühgeschichte der Bibliometrie“, S. 71.
[12] Ball, „Bibliometrische Dienstleistungen“, S. 556–557.

> their scientific performance in various ways. Such pressure is exercised not only by the administration but also by politics and the general public, but frequently even by the peers especially in academic appointment processes or for the allocation of research resources. This is somewhat understandable, because quantitative measures appear to be an easy way to determine whether the tax money is spent in a reasonable way. However, it is difficult to decide which way is reasonable, not only regarding the spending of the money but also regarding the measuring.[13]

Es ist und bleibt damit Aufgabe der Bibliometrie, sich dieser Kritik zu stellen und dem durch Einhaltung von wissenschaftlichen Standards sowie einer Qualitätssicherung der eigenen Methoden entgegenzuwirken. Das Bewusstsein dafür, dass alle bibliometrische Methoden ihre Grenzen haben und die Ergebnisse kritisch hinterfragt werden sollten, muss immer wieder gestärkt werden.
Über die Jahre hat sich das Thema Bibliometrie soweit etabliert, dass es nicht mehr nur ein reines Forschungsthema ist, sondern breite Anwendung in der Praxis von Bibliotheken und Forschungseinrichtungen findet. Trotzdem gibt es weiterhin umfangreiche Forschungsaktivitäten rund um das Thema, wozu auch die vorliegende Arbeit einen Beitrag liefern soll.

Geprägt wurde der Begriff „Bibliometrie“ 1969 von Alan Pritchard, der die bis dahin übliche Bezeichnung „*statistical bibliography*“ ablösen wollte:

> Therefore it is suggested that a better name for this subject [...] is bibliometrics, i.e. the application of mathematics and statistical methods to books and other media of communication.[14]

[13] Schreiber, „A skeptical view on the Hirsch index and its predictive power“, S. 1.
[14] Pritchard, „Statistical Bibliography or Bibliometrics“, S. 349.

Pritchard gab damit auch eine bis heute gültige Definition der Bibliometrie als „application of mathematics and statistical methods to books and other media of communication“[15]. Möglich und praktikabel geworden waren diese Methoden erst durch die Entwicklung des *Science Citation Index* duch Eugene Garfield seit Anfang der 1960er Jahre, der damit einen umfassenden interdisziplinären Index von Zitationen wissenschaftlicher Arbeiten geschaffen hat.[16] Ein wesentlicher Bestandteil der Bibliometrie ist damit die Zitatanalyse, also die Auswertung von Zitationen, dem „Bindeglied zwischen Publikationen“[17]. Die grundlegende Idee dahinter ist, dass ein Wissenschaftler in einem Artikel die Publikationen referenziert, welche für seine Arbeit relevant sind und eine inhaltliche Nähe haben. Daraus folgt, dass eine Publikation mit vielen Zitierungen große Resonanz in der Fachcommunity erfahren hat und damit vermutlich von größerer Bedeutung ist, als eine wenig zitierte. Bei diesen Schlussfolgerungen muss man jedoch Vorsicht walten lassen, denn „numerical indicators can reliably suggest only eminence but never worthlessness“[18]. Wesentliche Aufgabe der Bibliometrie ist es daher, geeignete Kennzahlen zu entwickeln, die die Rezeption von einzelnen Publikationen und Publikationssets zuverlässig beschreiben. Dafür wurden und werden immer bessere bibliometrische Indikatoren entwickelt, von denen im folgenden Abschnitt einige vorgestellt werden sollen. Keiner dieser Indikatoren eignet sich jedoch, um die Publikationstätigkeit einer untersuchten Einheit vollständig zu beschreiben. Daher wird es immer nötig sein, mehrere Kennzahlen geeignet zu kombinieren.[19]

1.2. Bibliometrische Kennzahlen

Im Laufe der Zeit wurde eine schier unüberschaubare Anzahl von zitationsbasierten bibliometrischen Indikatoren entwickelt, welche je nach Anwendungsgebiet unterschied-

[15] Pritchard, „Statistical Bibliography or Bibliometrics“, S. 349.
[16] Clarivate Analytics, *History of Citation Indexing*.
[17] Haustein und Tunger, „Sziento- und bibliometrische Verfahren“, S. 480.
[18] Braun und Schubert, „Dimensions of scientometric indicator datafiles“.
[19] Hornbostel, *Wissenschaftsindikatoren – Bewertungen in der Wissenschaft*, S. 326.

liche Facetten des wissenschaftlichen Outputs und dessen Wahrnehmung in der Fachcommunity beschreiben soll. Grob lassen sich diese Indikatoren in verschiedene Klassen einteilen – in eher einfache Zitationskennwerte, welche sich durch einfaches Zählen von Publikationen und Zitationen gewinnen lassen, in normalisierte Zitationsindikatoren, die versuchen, die Resonanz auf die untersuchten Publikationen relativ zu einer Vergleichsgruppe zu untersuchen und somit eine möglichst große Vergleichbarkeit der Kennzahlen anstreben, und zuletzt gewichtete Zitationskennzahlen, welche auf Netzwerktheoretischen Methoden beruhen.[20]

Das einfachste Beispiel aus der ersten Klasse von Zitationsindikatoren sind die Zitationsraten CPP, also die durchschnittliche Anzahl von Zitationen pro Publikation. Diese liefern eine größenunabhängige Kennzahl zum Vergleich von Autoren, Institutionen, Zeitschriften und vielem mehr. Deutlich weiter verbreitet zum Vergleich von Zeitschriften ist jedoch der *Journal Impact Factor*, der jährlich von Clarivate Analytics berechnet und in den *Journal Citations Reports* veröffentlicht wird. Mittlerweile wird dieser häufig dazu verwendet, die Qualität von Zeitschriften zu bewerten. Es gibt jedoch starke Kritik an diesem Vorgehen, besonders was die Art der Berechnung, mangelnde Reproduzierbarkeit und Aussagekraft betrifft.[21] Von der Popularität her vergleichbar mit dem *Journal Impact Factor* ist der 2005 von dem Physiker J.E. Hirsch vorgeschlagene h-Index, der die Rezeption von Publikationen einzelner Wissenschaftler beschreiben soll. Die Definition von Hirsch lautet:

> A scientist has index h if h of his or her N_p papers have at least h citations each and the other $(N_p - h)$ papers have $\leq h$ citations each.[22]

Ein Wissenschaftler mit 15 Publikationen und jeweils mindestens 15 Zitationen (nicht aber 16 Publikationen mit mindestens 16 Zitationen) hat demnach einen h-Index von 15. Auch an diesem Index gibt es viel Kritik[23], die Resonanz auf den Artikel von Hirsch

[20] Haustein und Tunger, „Sziento- und bibliometrische Verfahren“.
[21] Moed und Leeuwen, „Impact factors can mislead“.
[22] Hirsch, „An index to quantify an individual's scientific research output“, S. 16569.
[23] Waltman und Eck, „The inconsistency of the h-index“.

mit fast 4000 Zitationen in Scopus[24] war und ist jedoch gewaltig und zeigt damit, wie groß der Bedarf an einfachen bibliometrischen Metriken ist. In der vorliegenden Arbeit wird in Kapitel 3 die Fehleranfälligkeit des h-Index untersucht.

Alle der einfachen Zitationsmetriken haben den Nachteil, dass sie Aufgrund der unterschiedlichen Publikationsgewohnheiten der verschiedenen Fächer keine interdisziplinäre Vergleichbarkeit gewährleisten können. Mit Hilfe von normalisierten Zitationsmetriken versucht man dies zu umgehen und disziplinspezifische Unterschiede auszugleichen. Dabei gibt es verschiedene Möglichkeiten zur Normalisierung: Eine Möglichkeit, welche auch in dieser Arbeit untersucht wird, ist die Zitationsraten der untersuchten Dokumente ins Verhältnis zu setzen mit den durchschnittlichen Zitationsraten von allen Dokumenten in derselben Disziplin. Ein Wert von eins spiegelt dann eine durchschnittliche Wahrnehmung der Publikationen wieder. Werte größer als eins bedeuten überdurchschnittliche, Werte unter eins unterdurchschnittliche Wahrnehmung. Die Auswahl der Vergleichsdokumente kann zum Beispiel über eine fachliche Klassifizierung oder über die Zeitschrift, in der der jeweilige Artikel erschienen ist, erfolgen. Beide Fälle werden in Abschnitt 4 dieser Arbeit untersucht.
Eine andere Möglichkeit zur Normalisierung ist, die Länge von Referenzlisten bei der Zitationsanalyse zu berücksichtigen und eine Zitierung von einer Publikation mit wenig Referenzen höher zu werten, als eine mit vielen Referenzen. Auf diese Weise kann man auf die Definition der Referenzmenge verzichten.

Die dritte Klasse von Zitationsmetriken basiert auf netzwerktheoretischen Definitionen des Begriffes Zentralität. Dabei werden Zitationen durch relevante Publikationen als wichtiger angesehen, als die von weniger relevanten. Sie tragen somit wiederum stärker zu der Relevanz des zitierten Artikels bei. Dies liefert ein selbstkonsistent definiertes System von Gleichungen für die Relevanz aller Publikationen, welches iterativ gelöst

[24]Stand 06.20.2018

werden kann. Prominente Beispiele für diese Indikatoren sind der *Eigenfactor Score*[25] und der *SCImago Journal Rank*[26], die auf dem von Larry Page und Sergei Brin für die Internet-Suchmaschine Google entwickelten PageRank-Algorithmus[27] basieren, der in den Anfangsjahren von Google wesentlich zu dem Erfolg beigetragen hat. Metriken dieser Art werden in der vorliegenden Arbeit jedoch nicht untersucht.

Prinzipiell lassen sich alle dieser Indikatoren aus verschiedenen Datengrundlagen berechnen, wenn auch bei einigen Indikatoren die zu verwendenden Daten in der Definition vorgegeben sind.[28] Die jeweilige Datengrundlage sollte bei der Angabe der Werte des Indikators immer mit angegeben werde, da sich die Ergebnisse aufgrund unterschiedlicher Abdeckungsgrade der Datengrundlagen teilweise erheblich unterscheiden können. Üblicherweise für bibliometrische Analysen verwendete Zitationsdatenbanken sind *Web of Science*, *Scopus* und teilweise auch *Google Scholar*. Auf die jweiligen Vor- und Nachteile soll hier nicht weiter eingegangen werden, dazu gibt es bereits zahlreiche Untersuchungen.[29] [30]

Eine weitere bedeutende Quelle für bibliometrische Analysen bietet das seit Ende 2008 durch das Bundesministerium für Bildung und Forschung geförderte und aus sieben im Bereich der Bibliometrie aktiver Institutionen bestehende Kompetenzzentrum Bibliometrie: es hat „eine qualitätsgesicherte Inhouse Dateninfrastruktur, basierend auf den Datenbanken Scopus (Elsevier) und Web of Science (Thomson Reuters), aufgebaut und nutzt diese zur Entwicklung und Weiterentwicklung von Analysemethoden und Indikatoren.“[31] Die Mitglieder des Kompetenzzentrums haben damit einen direkten Datenbank-

[25] Bergstrom, „Eigenfactor: Measuring the value and prestige of scholarly journals“.

[26] Falagas u. a., „Comparison of SCImago journal rank indicator with journal impact factor“.

[27] Brin und Page, „The Anatomy of a Large-Scale Hypertextual Web Search Engine“.

[28] Dies gilt zum Beispiel für den *Journal Impact Factor*, der aus den Daten des *Web of Science* berechnet wird.

[29] Harzing und Alakangas, „Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison“.

[30] Li u. a., „Citation Analysis: Comparison of Web of Science®, Scopus™, SciFinder®, and Google Scholar“.

[31] Kompetenzzentrum Bibliometrie, *Über das Kompetenzzentrum Bibliometrie*.

zugriff auf die Zitationsdaten und können somit umfangreiche Analysen durchführen, welche allein mit den Datenbanken *Web of Science* oder *Scopus* durch technische oder lizenzrechtliche Einschränkungen nicht möglich wären. Ein weiterer Vorteil dieser Datenbank ist, dass die Inhalte „über eine Reihe von automatischen und semiautomatischen Prozeduren während der Ladeprozesse überprüft und Fehler entsprechend bekannter Muster korrigiert [werden]. Es werden insbesondere zahlreiche Vereinheitlichungen und Standardisierungen, z.B. von Zeitschriftennamen und Länderangaben, vorgenommen.“[32] Zudem gibt es eine Reihe vorberechneter Indikatoren, sowie eine implementierte Institutionencodierung deutscher Institutionen, die eine eindeutig Zuordnung von Publikation zur Institution ermöglichen. Die Bibliometriedatenbank wird jährlich aktualisiert und zu einem definierten Zeitpunkt eingefroren und archiviert, was die Reproduzierbarkeit der Analysen gewährleistet. Die in der vorliegenden Arbeit verwendeten Daten stammen von der Bibliometriedatenbank des Kompetenzzentrums.

1.3. Existierende Untersuchungen

Bislang gibt es relativ wenig Publikationen, welche sich mit Fehlerabschätzungen bei bibliometrischen Indikatoren beschäftigen. Das Bewusstsein für eine Notwendigkeit dieser Abschätzungen ist jedoch schon lange vorhanden, so schrieb Glänzel 2003 in einer Publikation:

> Bibliometric indicators are subject to a variety of errors (systematic, random errors and built-in data errors) that are usually not taken into account to the necessary extent when bibliometric data are applied in research evaluation.[33]

[32] Kompetenzzentrum Bibliometrie, *Dateninfrastruktur*.

[33] Glänzel und Debackere, „On the opportunities and limitations in using bibliometric indicators in a policy relevant context“, S. 227.

Trotzdem behandeln die meisten Untersuchungen, die sich mit diesem Thema beschäftigen schwerpunktmäßig den Bereich „Fehler in Zitationsdatenbanken“ und weniger die Auswirkungen davon auf bibliometrische Indikatoren. Häufig geschieht dies auch nur anhand von überschaubaren Fallbeispielen einzelner Wissenschaftler oder Institute, wie auch in der erwähnten Publikation von Glänzel.

Die bislang umfangreichste Arbeit zum Thema Fehler in Zitationsdatenbanken stammt wohl von Olensky. Hier wurden ausführlich verschiedene Fehler diskutiert und Fehlerwahrscheinlichkeiten für verschiedene Typen von Fehlern und Fächer angegeben.[34] Ähnliche Untersuchungen wurden auch von Franceschini durchgeführt.[35] Die hier gefundenen Fehlerquoten für fehlende Zitierungen liegen je nach Fach bei etwa 3% bis 12%.[36] Auswirkungen auf bibliometrische Analysen wurden in beiden Arbeiten nicht in größerem Maßstab untersucht.

Zum h-Index gibt es verschiedene Publikationen, welche seine Robustheit gegenüber Änderungen in der Datenbasis. Diese Publikationen verwenden teilweise mathematische Argumente[37], teilweise werden auch einzelne Beispiele zur Demonstration herangezogen[38].

Eine ausführliche Arbeit von Thelwall und Fairclough beschäftigt sich mit Fehleranalysen bei feldnormalisierten Indikatoren.[39] Dabei wurden jedoch keine realen Zitationszahlen verwendet, sondern künstlich aus einer diskretisierten Log-Normalverteilung statistisch simulierte Zitationsdaten. Für diese Daten wurden die Konfidenzintervalle zweier feldnormalisierter Indikatoren untersucht und miteinander verglichen.

[34] Olensky, „Data accuracy in bibliometric data sources and its impact on citation matching“.

[35] Franceschini, Maisano und Mastrogiacomo, „Empirical analysis and classification of database errors in Scopus and Web of Science“.

[36] Olensky, „Data accuracy in bibliometric data sources and its impact on citation matching“, S. 119.

[37] Courtault und Hayek, „On the Robustness of the h-index: a mathematical approach“.

[38] Vanclay, „On the robustness of the h-index“.

[39] Thelwall und Fairclough, „The accuracy of confidence intervals for field normalised indicators“.

— Hauptteil —

2. Vorgehensweise

Ziel dieser Arbeit ist die Analyse von den Auswirkungen fehlerhafter Datenbasen auf verschiedene bibliometrische Indikatoren. Grundsätzliche Vorgehensweise und technische Umsetzung der Untersuchungen sollen in diesem Abschnitt kurz erläutert werden. Weitere Details folgen dann in den jeweiligen Kapiteln 3 und 4 und im Anhang dieser Arbeit, wo einige Auszüge aus dem verwendeten Quellcode abgebildet sind.

Als Datengrundlage werden Auszüge aus der Bibliometriedatenbank WOS_B_2017 des Kompetenzzentrums Bibliometrie verwendet. Mit der frei verfügbaren Entwicklungsumgebung des Unternehmens Oracle für SQL werden die nötigen Rohdaten über SQL-Abfragen abgerufen und als csv-Dateien exportiert. Die jeweils relevanten Einträge in den Spalten der Datenexporte sind dabei die Identifikatoren der Publikationen, der *Subject Categories* und der Zeitschriften, die Publikationsjahre, Zitationszahlen und Dokumenttypen. Einen Auszug[1] der für diese Arbeit relevanten Tabellen aus dem Datenbankschema der Bibliometriedatenbank zeigt Abbildung 2.1. Die so erhaltenen csv-Dateien werden in Mathematica importiert, einer proprietären Software des Unternehmens Wolfram Research, welche ein Computeralgebrasystem, eine Numerik-Software, Visualisierungstools und eine eigene Programmiersprache enthält. Der Vorteil dieser Software sind die zahlreichen vorimplementierten Funktionen zur Durchführung der numerischen Rechnungen sowie zur Auswertung und Visualisierung der Daten.

Innerhalb von Mathematica werden die Daten aufbereitet und strukturiert sowie Pro-

[1] Das vollständige Datenbankschema ist deutlich komplexer und umfangreicher und ist leider nicht offen zugänglich.

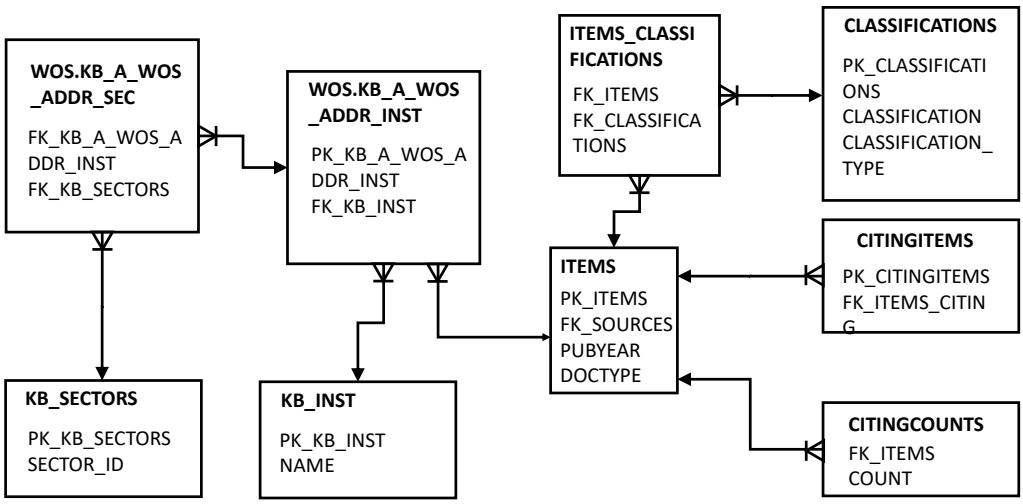

Abbildung 2.1.: Auszug aus dem Schema der relationalen Datenbank des Kompetenzzentrums Bibliometrie: Die in dieser Arbeit verwendeten Tabellen. Die mit KB bezeichneten Tabellen sind eigens vom Kompetenzzentrum erstellt worden und enthalten institutionsbereinigte Daten zu deutschen Einrichtungen.

gramme zur Berechnung der Zitationsmetriken entwickelt. Zudem werden Routinen programmiert, mit denen zufällige Fehlerkonfigurationen in die Datengrundlage gestreut werden können, genau so, dass eine Publikation oder Zitation gerade mit einer vorgegebenen Wahrscheinlichkeit gelöscht wird. Damit werden die fehlerbehafteten Zitationsmetriken für eine möglichst große Anzahl von Fehlerkonfigurationen berechnet und die Ergebnisse statistisch ausgewertet. Aus den Wahrscheinlichkeitsverteilungen dieser Zufallsvariablen lassen sich dann alle relevanten statistischen Größen wie Mittelwerte, Standardabweichungen und Konfidenzintervalle ableiten, aus denen wiederum Aussagen über die Fehleranfälligkeiten der betrachteten Zitationsmetriken abgeleitet werden können.

Alle Simulationen werden durchgeführt auf einem gewöhnlichen PC mit einem Prozessor der Grundtaktfrequenz 1,6 GHz und 4 Kernen sowie einem Arbeitsspeicher von 8 GB. Trotz Parallelisierung laufen die Simulationen insbesondere für die Rechnungen zu den normalisierten Indikatoren mehrere Tage. Nur so kann eine ausreichende Statistik erreicht werden.

3. Untersuchnung des h-Index

Der h-Index wurde 2005 von den argentinischen Physiker J.E. Hirsch als Indikator zur Beschreibung der Rezeption eines einzelnen Wissenschaftlers anhand dessen Publikationen vorgeschlagen.[1] und wird gemäß der Formel

$$h = \max\{r | c(r) \geq r\} \tag{3.1}$$

berechnet. Dabei ist $c(r)$ die Zitationszahl der r-ten Publikation, absteigend nach der jeweiligen Zitationszahl sortiert. Aufgrund der Einfachheit dieser Kennzahl hat sie sich bemerkenswert schnell verbreitet und findet häufig bei Bewertungsverfahren zu einzelnen Wissenschaftlern Anwendung. Trotzdem ist sie nicht unumstritten, so gibt es zahlreiche Kritikpunkte, welche nicht alle durch die ebenfalls vorhandenen Vorteile aufgewogen werden können.[2] Kennzahlen dieser Art wurden schon lange davor in anderen Bereichen verwendet.[3] [4] Eine häufig verwendete ebenfalls von Hirsch stammende graphische Interpretation ist in Abbildung 3.1 dargestellt. Aufgrund seiner besonderen Popularität und Einfachheit ist der h-Index ein naheliegender Kandidat als Untersuchungsobjekt auf seine Anfälligkeit gegenüber Fehlern in Zitationsdaten. Im Laufe der Analysen

[1] Hirsch, „An index to quantify an individual's scientific research output".

[2] Glänzel, „On the Opportunities and Limitations of the H-index".

[3] Ein relativ unbekanntes Beispiel ist die Anfang des 20. Jahrhunderts von dem britischen Physiker Sir Arthur Stanley Eddington vorgeschlagene *Eddington number of cycling*, ein Index der die Leistungen von Langstreckenradfahrern beschreiben soll: Ein Radfahrer hat eine Eddingtonzahl von E, wenn er E Tage mehr als E Meilen gefahren ist. Die Zahl E ist damit dimensionslos, jedoch davon anhängig, welche Längeneinheit bei der Definition von E verwendet wird. Eddington selber hatte einen Wert von $E[\text{Meilen}] = 84$. Siehe „Physics and sport", S. 15

[4] Ramsay, „Cycling record".

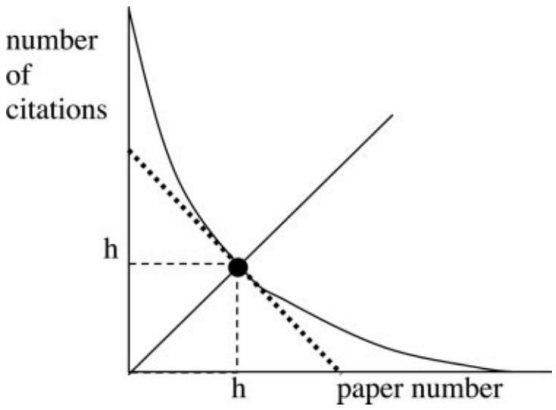

Abbildung 3.1.: Graphische Darstellung der Bedeutung des h-Index: Die Anzahl der Zitierungen wird gegen die Position des zugehörigen Artikels im Ranking abfallend nach der Anzahl der Zitationen aufgetragen (abfallende Kurve). Der Schnittpunkt der Winkelhalbierenden mit der Kurve ist dann der h-Index. [Entnommen der Originalpublikation: (Hirsch, „An index to quantify an individual's scientific research output")]

lässt sich dabei bereits einiges lernen, was anschließend bei der Untersuchung komplexerer Indikatoren Anwendung finden kann. Durch die einfache Art der Berechnung eignet er sich besonders zur Entwicklung Methodik der bibliometrischen Fehleranalyse.

3.1. Methodik

In diesem Abschnitt der Arbeit sollen die Einflüsse von zwei verschiedenen Arten von Fehlern auf den h-Index verglichen werden:

- Zunächst wird der Fall untersucht, bei dem ganze Publikationen des untersuchten Autors in der Datengrundlage fehlen. Bei der Berechnung des h-Index eines Autors werden also nur M_{err} Publikationen statt der eigentlichen Anzahl berücksichtigt. Mit $M_{\text{err}} \leq M$ folgt daraus dann zwingend ein h-Index $h_{\text{err}} \leq h_0$, wobei h_0 den ungestörten h-Index bezeichnet – berechnet aus den M Publikationen – und h_{err} den gestörten h-Index – berechnet aus den M_{err} Publikationen. Dieser Fehler kommt

bei realistischen Analysen zu real existierenden Personen vor, wenn eine Publikation fälschlicherweise einem anderen Autor zugeordnet wurde oder eine Publikation schlicht in der Datengrundlage fehlt. Der umgekehrte Fall, dass dem untersuchten Autor eine Publikation eines anderen zugeordnet wurde, soll hier nicht weiter behandelt werden. Zudem wird der Effekt vernachlässigt, der dadurch entsteht, dass für jede Publikation, die aus der Datenbasis entfernt wird, eigentlich auch die Zitationszahlen der durch diese Publikation zitierten Artikel aus der Datenbasis um eins reduziert werden müssten. Die Umsetzung davon wäre jedoch sehr aufwändig, weshalb darauf verzichtet wird.

- Im zweiten Fall werden statt ganzer Publikationen bei der Berechnung des h-Index nur einzelne Zitationen der fehlerbehafteten Publikationen entfernt. Diese Art von Fehler kommt in den üblichen für bibliometrische Analysen verwendeten Datenbanken recht häufig vor, wenn zwar die betrachtete Publikation korrekt eingetragen und dem richtigen Autor zugeordnet ist, dagegen aber eine der zitierenden Publikationen nicht in der Datenbank vorhanden ist oder die Verlinkung der beiden Publikationen fehlt.

In beiden Fällen erhält man zwei verschiedene Grundgesamtheiten, aus denen anschließend die h-Indizes bestimmt werden: Zunächst die ungestörte, welche den Daten der Datenbank des Kompetenzzentrums entnommen wird, und dann die fehlerbehaftete, welche statistisch generierte Fehler mit kontrollierter Fehlerwahrscheinlichkeit enthält. Erstere wird im Rahmen dieser Arbeit als exakt angenommen – auch wenn hier selbstverständlich Fehler enthalten sein werden – und zur Bestimmung der exakten h-Indizes h_0 verwendet. Aus letzterer werden die fehlerbehafteten h-Indizes h_{err} berechnet und diese mit h_0 verglichen. Als Autoren werden dabei keine realen Personen verwendet, sondern zufällige Untermengen der Grundgesamtheit mit dem Umfang M verwendet. Dies hat den Vorteil, dass man nicht auf Autorenprofile zurückgreifen muss.
Für beide Fehlerarten gibt es verschiedene Möglichkeiten zur Implementierung. Davon werden in dieser Arbeit zwei näher betrachtet: Zunächst wird eine feste Anzahl von Pu-

blikationen der Grundgesamtheit verändert, genau so, dass dies der gewünschten Fehlerquote entspricht: Bei den Fehlern erster Art werden also $p \cdot N_0$ Publikationen gelöscht, beziehungsweise die Zitationszahl dieser Publikationen auf null gesetzt, damit sie nicht zum h-Index einer Person beitragen können. N_0 ist dabei die Anzahl von Publikationen in der Grundgesamtheit, also aller untersuchter Publikationen. Für die Untersuchung der Fehler zweiter Art werden dementsprechend $p \cdot N_0 \cdot \mathrm{CPP}$ Zitationen zufällig entfernt, wobei CPP die durchschnittliche Zitationszahl der Publikationen in der Grundgesamtheit bezeichnet.
Eine andere Möglichkeit zur Fehlerimplementierung ist, nicht die Gesamtzahl der Fehler fest vorzugeben, sondern nur die Wahrscheinlichkeit p, mit der eine Publikation (oder Zitation) fehlt. Wie sich später zeigt, ist diese Art der Implementierung deutlich einfacher umzusetzen und hat zudem den Vorteil, dass die Rechenzeit für vergleichbare Genauigkeiten deutlich verkürzt werden kann. Mittelt man dabei über eine hohe Anzahl von Stichproben, so sind die Resultate der beiden Methoden identisch. Sie werden in Abschnitt 3.2.3 auf Basis der Datengrundlage der Publikationen der Physik miteinander verglichen.

3.2. Untersuchung der Publikationen der Physik

Als Datengrundlage für die vorliegende Untersuchung werden zunächst alle wissenschaftlichen Publikationen deutscher Universitäten mit der fachlichen Zuordnung zur Physik untersucht. Dazu wird die Institutionsbereinigung der Datenbank des Kompetenzzentrums verwendet, die zu jeder Publikation aus Deutschland verzeichnet, von welchem Institutionstyp die Verfasser kommen. Für die Festlegung auf Publikationen aus der Physik werden die *Subject Categories* des *Web of Science* verwendet.[5] Für die Physik werden die in Tabelle 3.1 angegebenen *Subject Categories* verwendet. Zudem wird die Datengrundlage auf die in der Physik gängigen Publikationsformen „Article“ und

[5] Clarivate Analytics, *Web of Science Core Collection Help*.

Identifikator	81	85	136	137	202
Beschreibung	Physics, Mathematical	Physics, Fluids & Plasmas	Physics, Particles & Fields	Physics, Nuclear	Physics
Identifikator	265	294	355		
Beschreibung	Physics, Condensed Matter	Physics, Atomic, Molecular & Chemical	Physics, Multidisciplinary		

Tabelle 3.1.: In diesem Abschnitt verwendete *Subject Categories* des *Web of Science* zur Festlegung der Publikationen in der Physik.

„Review“ eingeschränkt, über welche hier der Großteil der wissenschaftlichen Kommunikation abläuft und auf die fast alle Zitationen in der Physik entfallen. Die Physik eignet sich in besonderem Maße für Untersuchungen dieser Art, da es sich hierbei um ein klassischen *Zeitschriftenfach* handelt und hier der Großteil der wissenschaftlichen Publikationen in Form von Zeitschriftenartikeln veröffentlicht wird. Zudem ist die Abdeckung der Physik im *Web of Science* sehr hoch. Zugleich ist der Umfang der Datengrundlage noch überschaubar, was den zeitlichen Aufwand für die numerischen Simulationen im Rahmen hält. Auch Hirsch verwendete bekannte Wissenschaftler aus der Physik für seine Grundlegende Untersuchung[6].

3.2.1. Voruntersuchung der Datengrundlage

Bevor die eigentliche Untersuchung der Auswirkungen der verschiedenen Fehler auf die h-Indizes in der Physik begonnen wird, sollen einige Eigenschaften der Datengrundlage

[6]Hirsch, „An index to quantify an individual's scientific research output“.

analysiert und vorgestellt werden. Dies erleichtert die Einordnung der späteren Ergebnisse und besonders den Vergleich verschiedener Fächer in Abschnitt 3.3 deutlich.
Insgesamt erhält man durch die in der Einleitung zu Abschnitt 3.2 definierte Grundgesamtheit 254832 Publikationen mit zusammen 6094029 Zitationen und somit durchschnittlich 23,91 Zitationen pro Publikation. Die Anzahl nicht zitierter Publikationen liegt bei 21717, was einem Anteil von 8,5% entspricht.
Die Verteilung der Zitationen ist als Histogramm in Abbildung 3.2 dargestellt. Auf den

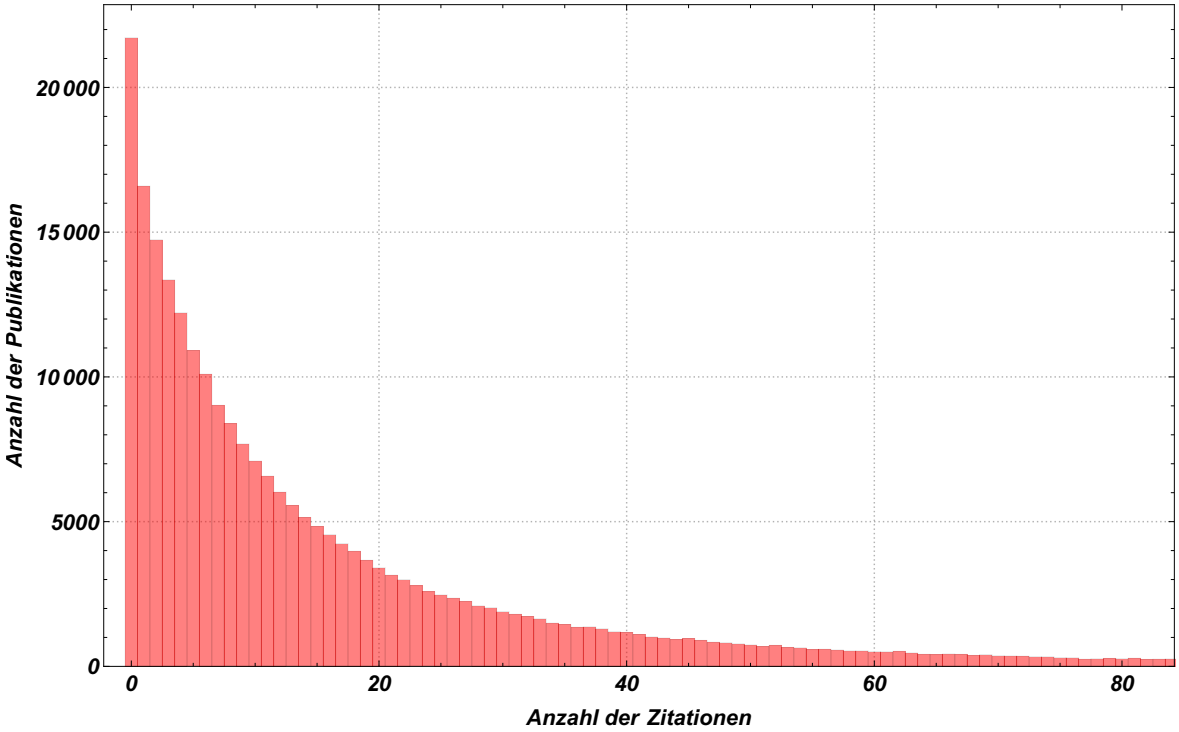

Abbildung 3.2.: Verteilung der Zitationen auf die deutschen Publikationen aus der Physik

ersten Blick scheint das analog zum Zipfschen Gesetz[7], welches häufig die Verteilung von bestimmten in eine Rangfolge gebrachte Größen beschreibt, einer Pareto-Verteilung und somit einem Potenzgesetz zu genügen. Dabei muss jedoch berücksichtigt werden, dass in der Arbeit von Zipf die Häufigkeit gegen den Rang aufgetragen wurde, hier jedoch die Häufigkeit gegen die entsprechende Zitationszahl. Die doppellogarithmische Auftragung in Abbildung 3.3 zeigt, dass es hier deutliche Abweichungen von dem einfachen Potenzge-

[7]Zipf, „The Meaning-Frequency Relationship of Words“.

setz $\sim n^{-s}$ gibt, welches hier als einfache Gerade mit der Steigung $-s$ erscheinen würde. Vielmehr lassen sich die Daten sehr genau mit dem Zipf-Mandelbrot-Gesetz[8] beschreiben, einer vom französischen Mathematiker Mandelbrot vorgeschlagenen Verallgemeinerung des Zipfschen Gesetzes mit der Wahrscheinlichkeitsverteilung

$$f(n; a, q, s) = \frac{a}{(n+q)^s}. \tag{3.2}$$

Die Normierungskonstante a ergibt sich dabei aus der Normierungsbedingung

$$N = \sum_n f(n; a, q, s) \Rightarrow a = \frac{N_0}{\sum_n \frac{1}{(n+q)^s}}, \tag{3.3}$$

wobei N_0 die Gesamtzahl der Publikationen ist, und q und s zwei freie Parameter, die an die gegebenen Daten angepasst werden müssen. Abbildung 3.3 zeigt die Übereinstimmung zwischen numerischen Daten und der angefitteten Verteilung. Die Beobachtung, dass die Verteilung der Zitationszahlen physikalischer Publikationen durch das Zipf-Mandelbrot-Gesetz beschrieben wird, wurde bereits in einer über den Preprint-Server arXiv.org veröffentlichten Arbeit von Silagadze gefunden.[9] Dabei wurden allerdings nur einzelne Wissenschaftler und Untermengen der Physik deutlich kleineren Umfangs untersucht. 2012 wurde von Egghe das Verhalten der h-Indizes für Zitationen mit dieser Verteilung analysiert. Dabei wurde jedoch nur der Wert $q = 1$ betrachtet[10].

Auf Basis dieser Publikationen sollen nun die h-Indizes fiktiver Autoren bestimmt und die zugehörige Verteilung untersucht werden. Dazu werden zufällige Publikationssets des Umfangs M generiert, indem aus allen Publikationen zufällig M Publikationen ausgewählt werden.[11] Für jeden dieser fiktiven Autoren wird dann der h-Index berechnet.

[8] Mandelbrot, „An informational theory of the statistical structure of language“, S. 491.

[9] Silagadze, „Citations and the Zipf-Mandelbrot's law“.

[10] Egghe und Rousseau, „The Hirsch index of a shifted Lotka function and its relation with the impact factor“.

[11] Dieser Prozess kann natürlich noch deutlich verfeinert werden, wenn die Erscheinungsjahre der gewählten Publikationen berücksichtigt werden und für jeden fiktiven Forschenden eine aktive Zeit und eine durchschnittliche Anzahl von Publikationen pro Jahr angenommen würde. Auf diese Weise kämen jedoch neben der Anzahl an Publikationen eines Autors mindestens drei weitere Parameter ins Spiel, was die Analyse deutlich aufwändiger machen würde. In dieser Arbeit soll dies daher noch nicht berücksichtigt werden.

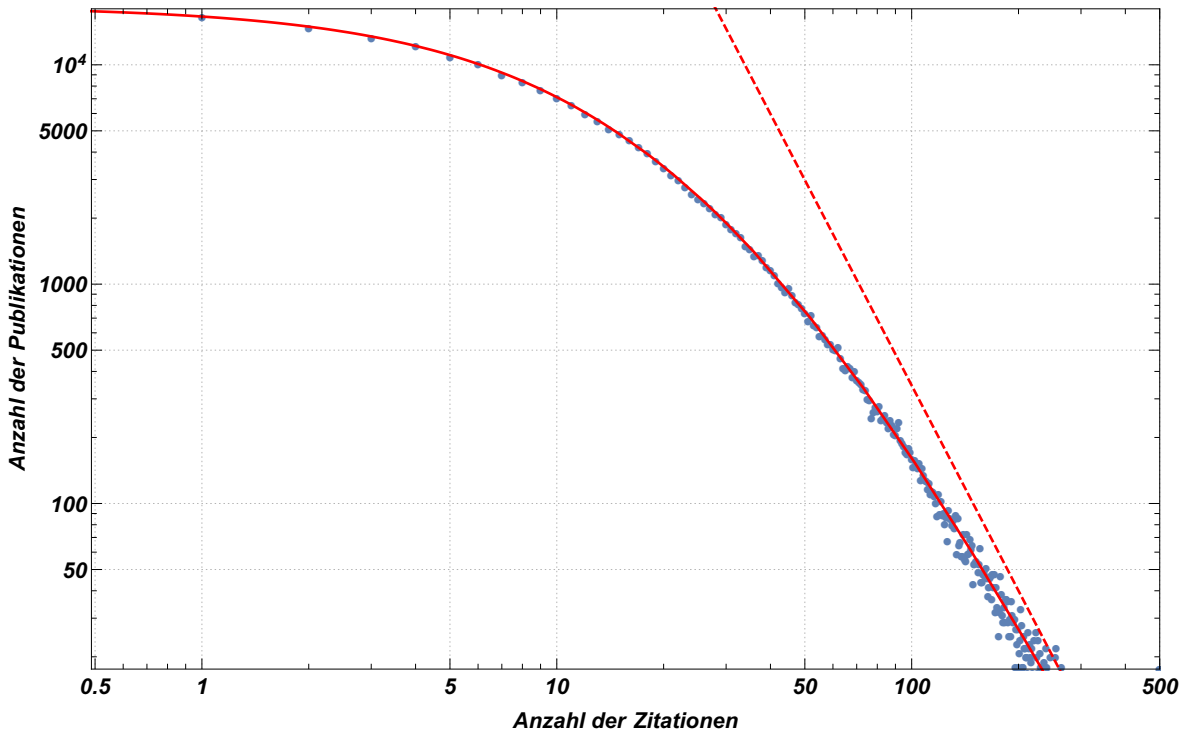

Abbildung 3.3.: Verteilung der Zitationen auf die deutschen Publikationen aus der Physik in doppellogarithmischer Auftragung. Die durgezogene Linie zeigt den *Fit* einer Zipf-Mandelbrot-Verteilung mit den Parametern $a = 573187453{,}8$, $s = 3{,}11$ und $q = 27{,}79$, die strichlierte Linie die gewöhnliche Zipf-Verteilung mit $a = 573187453{,}8$ und $s = 3{,}11$, die die Daten für große n näherungsweise beschreibt (siehe Haupttext).

Wiederholt man dies M_M mal für jeweils neu ausgewählte zufällige Publikationssets, so konvergiert die Verteilung der h-Indizes im Limes $M_M \to \infty$ statistisch gegen einen Grenzwert, die (diskrete) Wahrscheinlichkeitsverteilung des h-Index.

Diese Wahrscheinlichkeitsverteilungen sind in Abbildung 3.4 für $M = 25, 50$ und 100 für jeweils $M_M = 100000$ Konfigurationen dargestellt. Je größer die Anzahl M der Publikationen eines Autors, desto weiter verschiebt sich die Verteilung Richtung größerer h-Indizes und die Breite der Verteilung nimmt zu: Die Mediane liegen jeweils bei 11, 17 beziehungsweise 25, die Standardabweichungen bei 1,56, 1,86 beziehungsweise 2,18.

Interessant ist auch die Wahrscheinlichkeitsverteilung, welcher die h-Indizes für ein festes, großes M genügen. Abbildung 3.5 zeigt die numerischen Daten für $M = 10^3$ und $M_M = 10^6$ in logarithmischer Auftragung. Offensichtlich ist die Verteilung schief und

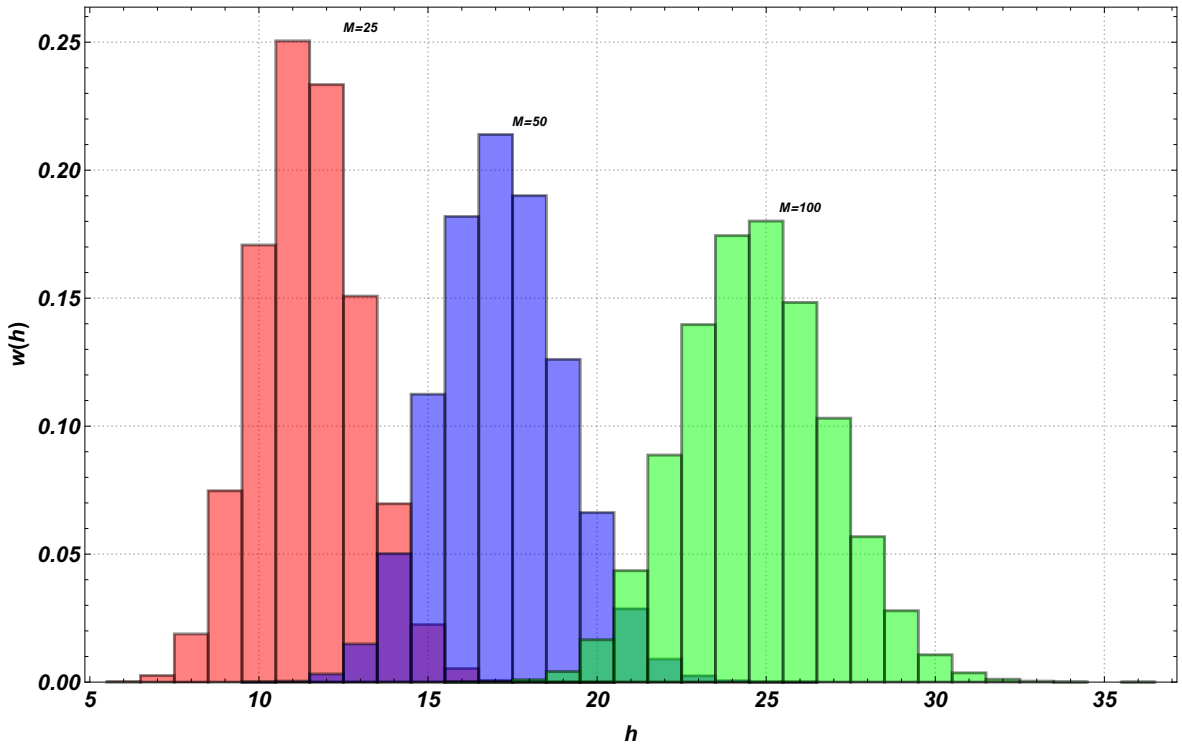

Abbildung 3.4.: Verteilung der h-Indizes der Physik für $M = 25, 50, 100$ Publikationen. Gemittelt wird über $M_M = 100000$ zufällig gewählte Konfigurationen.

lässt sich somit nicht durch eine Normalverteilung modellieren. Der als rote durchgezogene Linie dargestellte Fit einer logarithmischen Normalverteilung

$$w(h) = ae^{-b\ln(ch)} \tag{3.4}$$

beschreibt mit den Koeffizienten $a = 0{,}125$, $b = 232{,}72$ und $c = 0,0146$ die Daten hingegen sehr genau.

Im nächsten Abschnitt sollen in diese Datengrundlage statistische Fehler eingebaut und der Einfluss dieser Fehler auf die sich aus den Daten ergebenden h-Indizes untersucht werden. Dies geschieht zunächst wieder nur am Beispiel der Physik, bevor in Abschnitt 3.3 auch andere Fächer untersucht werden.

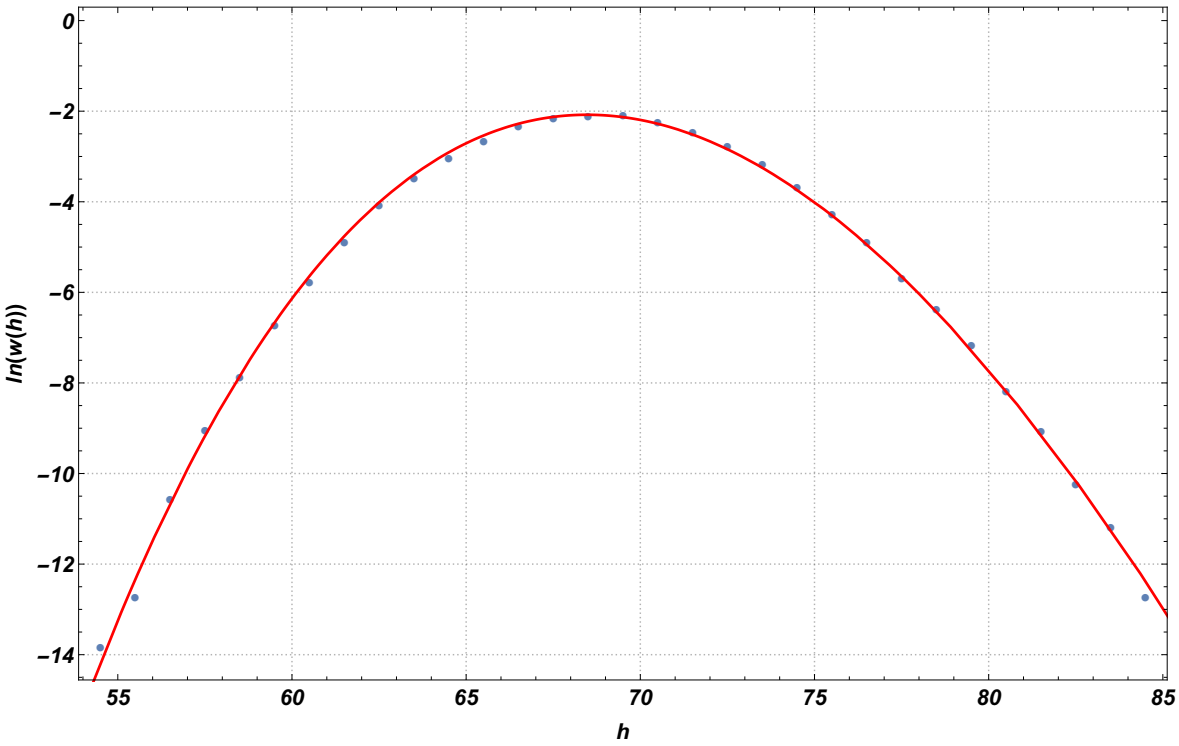

Abbildung 3.5.: Verteilung der h-Indizes der Physik für $M = 10^3$ in logarithmischer Auftragung. Gemittelt wird über $M_M = 10^6$ zufällig gewählte Konfigurationen. Die Daten werden sehr gut durch eine Log-Normalverteilung beschrieben und sind somit schief verteilt.

3.2.2. Fehlende Publikationen

Wie bereits in Abschnitt 3.1 erwähnt, gibt es verschiedene Arten von Fehlern, welche statistisch in die Datengrundlage eingebaut werden können. Zunächst sollen Fehler untersucht werden, bei der einzelne Publikationen fehlen. Auch dazu gibt es verschiedene Möglichkeiten der Implementierung, deren Vor- und Nachteile im Folgenden näher betrachtet werden sollen.

Will man eine fest vorgegebene Anzahl von Fehlern N_{err} in die Grundgesamtheit implementieren, so müssen bei großen Werten von N_{err} sehr viele Fehlerkonfigurationen zwischengespeichert werden, was den numerischen Aufwand sehr hoch macht. Die meisten dieser eingebauten fehlerhaften Publikationen werden wegen $M \ll N_0$ bei der Berechnung der gestörten h-Indizes aber nicht benötigt. Daher ist es numerisch deutlich sinnvoller, erst die M Publikationen zufällig zu bestimmen und anschließend die Feh-

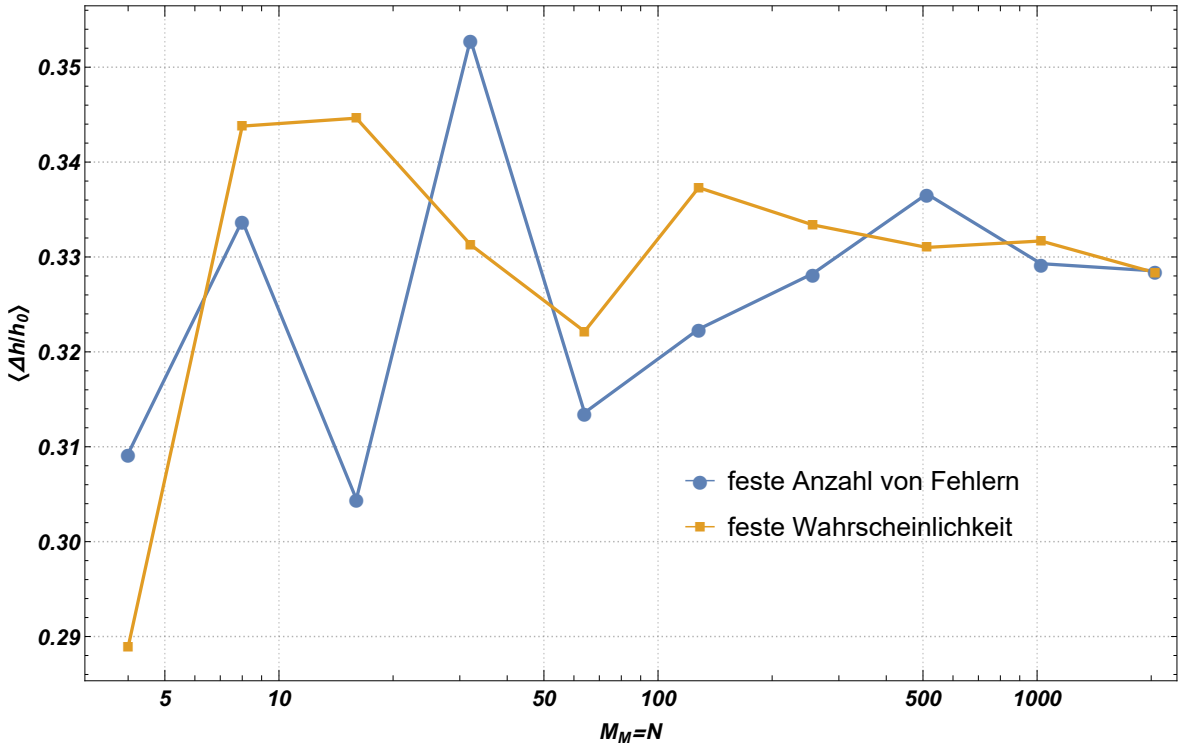

Abbildung 3.6.: Vergleich der beiden in dieser Arbeit vorgestellten Methoden zur Fehlerimplementierung: Dargestellt ist der gemittelte statistische Fehler des h-Index gegen die Anzahl der Konfigurationen $M_M = N$, über die bei festem $p = 0{,}5$ und $M = 50$ gemittelt wird. Die blauen Daten zeigen den Fall, bei dem eine Feste Anzahl von Fehlern in die Grundgesamtheit implementiert werden, die gelben Daten den Fall fester Wahrscheinlichkeiten für jede Publikation. Die Daten konvergieren für eine große Anzahl von Konfigurationen gegen einen gemeinsamen Grenzwert.

ler auf diese zu verteilen. Dabei ist a priori aber nicht klar, wie viele Elemente dieses Datensatzes tatsächlich Fehler aufweisen: im Extremfall kann es sein, dass alle M ausgewählten Publikationen fehlerhaft sind, oder aber, dass keine Fehler aufweist. Es ist jedoch möglich, die Wahrscheinlichkeit anzugeben, mit der genau $0 \leq m \leq M$ der ausgewählten Publikationen fehlerbehaftet sind. In der Wahrscheinlichkeitstheorie entspricht das einem Urnenmodell, bei dem aus N_0 Kugeln, von denen N_{err} weiß (fehlerbehaftet) und $N_0 - N_{\text{err}}$ schwarz (korrekt) sind, zufällig genau M ausgewählt werden. Dabei werden die Kugeln nach dem Ziehen nicht zurückgelegt. Die Anzahl k weißer Kugeln unter den M gezogenen genügt dann der hypergeometrischen Verteilung und für

die Wahrscheinlichkeit genau k weiße Kugeln zu ziehen gilt dann[12]

$$w(k) = \frac{\binom{N_{\text{err}}}{k}\binom{N_0 - N_{\text{err}}}{M-k}}{\binom{N_0}{M}}. \tag{3.5}$$

Dabei bezeichnet $\binom{n}{m} = \frac{n!}{m!(n-m)!}$ den Binomialkoeffizienten. Das Verhältnis $p = N_{\text{err}}/N_0$ ist die Wahrscheinlichkeit, dass eine beliebige Publikation aus der Grundgesamtheit einen Fehler aufweist.

Mit diesem Wissen genügt es also, in jede Konfiguration von M gezogenen Publikationen mit der Wahrscheinlichkeit $w(k)$ genau k Fehler einzubauen. Diese Vorgehensweise reduziert den Rechen- und Speicheraufwand der Simulation erheblich.

Die andere in dieser Arbeit verwendete Möglichkeit zur Implementierung der Fehler in die Datengrundlage ist deutlich einfacher umzusetzen: hierbei wird nicht die Anzahl der fehlerhaften Einträge in der Datengrundlage fest vorgegeben, sondern für jede der M zufällig ausgewählten Publikationen unabhängig voneinander entschieden, ob diese fehlerhaft ist oder nicht. Dabei wird jeder Eintrag mit der Wahrscheinlichkeit p gelöscht und mit der Wahrscheinlichkeit $1 - p$ unverändert übernommen.

Mit beiden Methoden können nun zufällig Fehlerkonfigurationen erzeugt werden. Dabei genügt es natürlich nicht, nur eine Konfiguration zu erzeugen, sondern es muss wieder – zusätzlich zu den in Abschnitt 3.2.1 eingeführten Mittelungen über M_M Publikationskonfigurationen – über $N \gg 1$ Fehlerkonfigurationen gemittelt werden. Insgesamt werden also $N \cdot M_M$ fehlerbehaftete h-Indizes berechnet und anschließend gemittelt.

Abbildung 3.6 zeigt exemplarisch für $M = 50$ und $p = 1/2$, wie die beiden Methoden für eine große Anzahl von Fehler- und Publikationskonfigurationen $M_M = N = 2^1, 2^2, \ldots, 2^{11}$ gegen den selben statistischen Grenzwert konvergieren. Aufgetragen ist hier die gemittelte relative Abweichung des Fehlerbehafteten h-Index h_{err} von dem ungestörten h_0:

$$\left\langle \frac{\Delta h}{h_0} \right\rangle = \frac{1}{M_M \cdot N} \sum_{\{M\},\{N\}} \frac{h_{\text{err}}(\{M\},\{N\}) - h_0(\{M\})}{h_0(\{M\})}. \tag{3.6}$$

Summiert wird dabei über alle Publikationskonfigurationen $\{M\}$ und alle Fehlerkonfigurationen $\{N\}$, wobei h_0 nur von den Publikationskonfigurationen abhängig ist, h_{err}

[12] Weisstein, *Hypergeometric Distribution. From MathWorld–A Wolfram Web Resource.*

jedoch von beiden.
Im statistischen Limes großer Konfigurationszahlen ist es also unerheblich, welche Methode zur Fehlerimplementierung verwendet wird. Daher wird im Folgenden die Methode fester Wahrscheinlichkeiten p pro Publikation verwendet, da diese einfacher zu implementieren und der Rechenaufwand geringer ist.

Die wesentlichen Ergebnisse dieses Abschnittes sind in Abbildung 3.7 grafisch dargestellt. Gezeigt sind die relativen Fehler der h-Indizes in Abhängigkeit der Fehlerwahrscheinlichkeit p einzelner Publikationen für verschiedene $M = 10, 25, 50, 75$ und 100. Gemittelt wird dabei jeweils über $M_M = N = 500$ Konfigurationen. An den relativ glatten Kurven erkennt man, dass Konfigurationszahlen in dieser Größenordnung bereits ausreichen, um brauchbare qualitative Ergebnisse zu bekommen. Interessant ist, dass alle Kurven unterhalb der schwarz strichliert eingezeichneten Winkelhalbierenden $\langle \Delta h/h_0 \rangle = p$ liegen. Dies bedeutet, dass der h-Index robust gegen die hier betrachteten Fehler ist, beziehungsweise Fehler in der Datenbasis bei der Berechnung des h-Index nicht zusätzlich verstärkt werden. Besonders für kleine Werte von p ist dieser Unterschied stark ausgeprägt, da die simulierten Kurven eine Steigung von deutlich unter eins aufweisen. Offensichtlich muss gelten, dass bei $p = 0$ auch der gemittelte relative Fehler im h-Index verschwindet und bei $p = 1$ auch der gemittelte relative Fehler gegen eins geht. Diese Eigenschaften weisen alle Kurven in Abbildung 3.7 auf. Zudem erkennt man, dass die Fehler mit steigendem M (und somit auch mit steigendem h-Index) geringer werden.

In der Realität sind Fehler in Zitationsdaten schief verteilt. Dies bedeutet, dass selbst relativ geringe absolute Fehlerhäufigkeiten dazu führen können, dass in einzelnen Stichproben sehr viele falsche Einträge auftreten. Dies ist besonders bei häufig auftretenden Autorennamen der Fall, da hier eine eindeutige Zuordnung von Publikationen zu Autoren sehr schwierig ist. Daher ist es notwendig, die gesamte Kurve in Abbildung 3.7 zu

berücksichtigen und sich nicht nur auf kleine Werte von p zu beschränken.[13]
Zudem treten in realen Systemen auch Fehler auf, die dafür sorgen, dass der fehlerbehaftete h-Index *größer* als der richtige sein kann. Dies wäre zum Beispiel der Fall, wenn dem betrachteten Autor fälschlicherweise eine Publikation zugeordnet würde, die eigentlich nicht von ihm stammt. Da solche Phänomene in den hier durchgeführten Untersuchungen nicht berücksichtigt werden, sind die hier gefundenen Fehler tendenziell größer als in der Realität, da sich dann Fehler verschiedener Art teilweise aufheben können.

3.2.3. Fehlende Zitationen

Nachdem im letzten Abschnitt untersucht wurde, wie sich das Fehlen ganzer Publikationen auf daraus resultierende h-Indizes auswirkt, soll nun analysiert werden, was sich ändert, wenn nicht ganze Publikationen, sondern nur einzelne Zitationen in der Datengrundlage fehlen. Diese Analysen werden wieder am Beispiel der deutschen Publikationen in der Physik durchgeführt, damit die Ergebnisse vergleichbar sind. Dabei hat man wie auch schon im vorherigen Abschnitt zwei Möglichkeiten zur Fehlerimplementierung. Der höhere Aufwand bei der Implementierung einer festen Anzahl von fehlenden Zitationen in der Grundgesamtheit macht sich durch die deutlich größere Gesamtzahl dieser (mit der durchschnittlichen Zitationszahl in der Physik von 23,91 als Faktor) hierbei noch stärker bemerkbar. Insbesondere steigt der Rechenaufwand stark an, weshalb hier nur der Fall untersucht wird, bei dem für jede einzelne Zitation in der Grundgesamtheit eine feste Wahrscheinlichkeit p angenommen wird, mit der diese fehlt. Für eine Publikation mit n Zitationen fehlen also mit der Wahrscheinlichkeit $\binom{n}{i}p^i(1-p)^{n-i}$ genau $0 \leq i \leq n$ dieser n Zitationen.

Abbildung 3.8 zeigt die Ergebnisse für den relativen gemittelten Fehler für verschiedene

[13] Bornmann und Leydesdorff, „Skewness of citation impact data and covariates of citation distributions: A large-scale empirical analysis based on Web of Science data“.

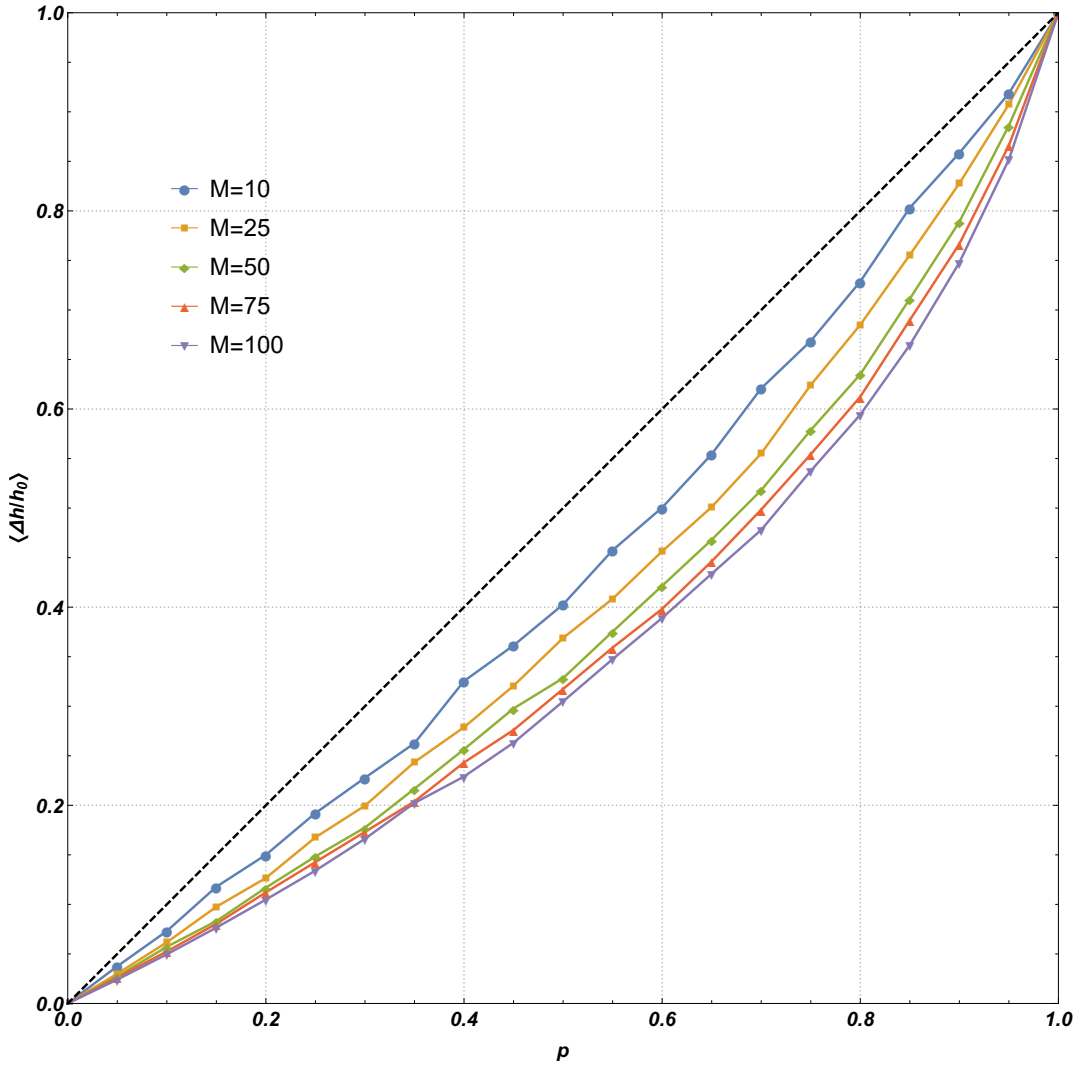

Abbildung 3.7.: Relativer Fehler der h-Indizes in Abhängigkeit von der Fehlerwahrscheinlichkeit p einzelner Publikationen für verschiedene Werte von M.

feste Werte von M mit $M_M = N = 500$ Publikations- und Fehlerkonfigurationen in Abhängigkeit von der Wahrscheinlichkeit p, dass eine beliebig gewählte Zitation fehlt. Die Kurven verlaufen ähnlich wie im vorherigen Abschnitt: für $p = 0$ beginnen sie alle bei null mit einer Steigung kleiner als eins und liegen im gesamten Intervall unterhalb der Winkelhalbierenden. Dies bedeutet, dass auch hier der h-Index robust gegen die eingebauten Fehler ist und diese abschwächt. Im Limes $p \to 1$ geht der Fehler wie zu erwarten wieder gegen eins. Es gibt jedoch zwei wesentliche Unterschiede gegenüber dem Fall fehlender Publikationen:

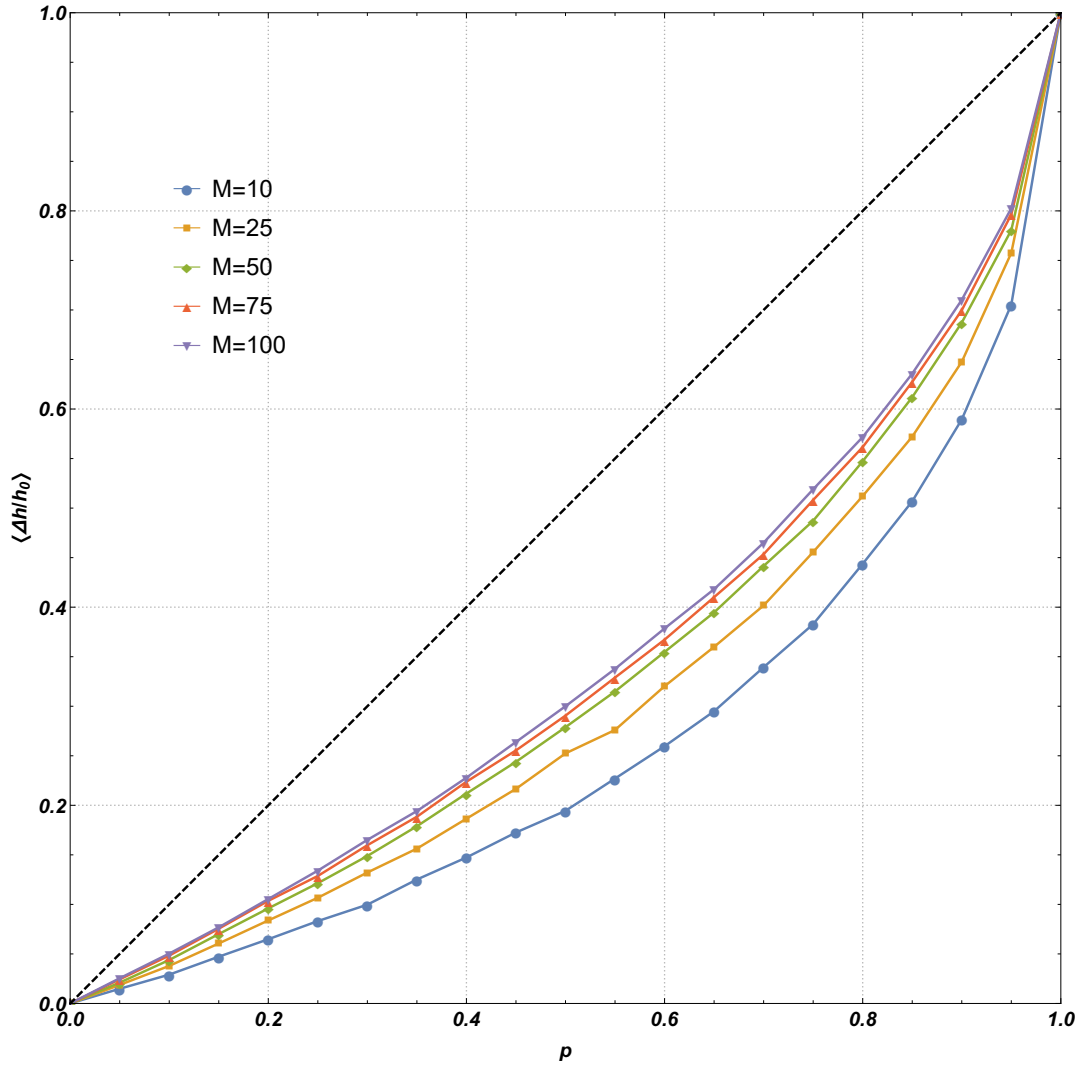

Abbildung 3.8.: Relativer Fehler der h-Indizes in Abhängigkeit von der Fehlerwahrscheinlichkeit p einzelner Zitationen für verschiedene Werte von M.

- Zum einen sind relativen gemittelten Fehler hier signifikant geringer als im vorangegangenen Abschnitt. Alle Kurven in Abbildung 3.8 liegen deutlich unter denen von Abbildung 3.7. Abbildung 3.9 zeigt die beiden Abbildungen zusammen in einem Diagramm. Für kleine Werte von M ist dieser Effekt deutlich stärker ausgeprägt als für große M. Bei $M = 100$ sind die beiden Kurven nahezu identisch.

- Zum anderen steigen die relativen gemittelten Fehler bei festem p mit wachsendem M an, statt wie in Abschnitt 3.2.2 geringer zu werden.

Während man den ersten Unterschied intuitiv erwarten würde ist die zweite Beobachtung eher überraschend. Hierfür fehlt bislang eine einfache Erklärung.

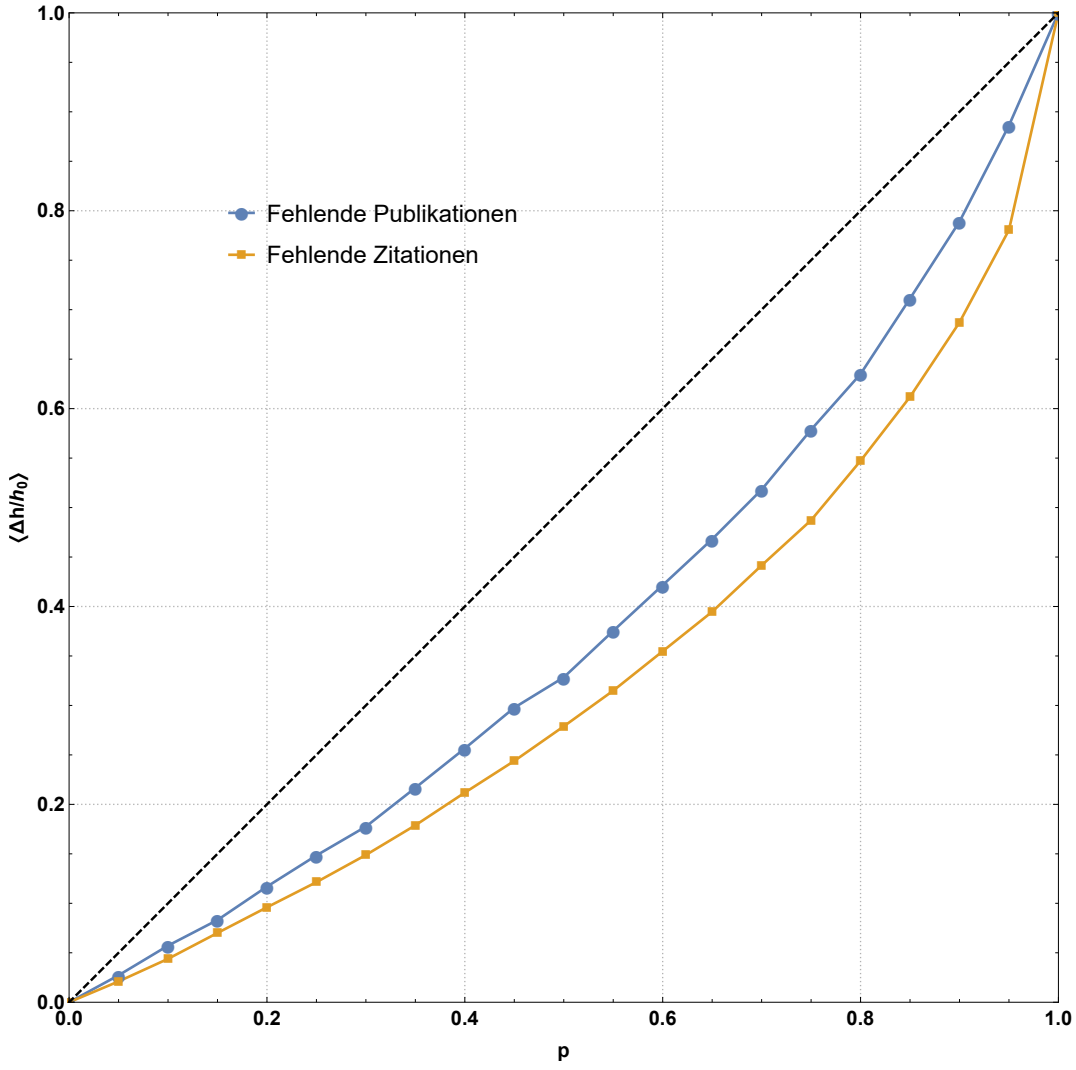

Abbildung 3.9.: Vergleich der Auswirkungen von fehlenden Publikationen und fehlenden Zitationen anhand der Physik für $M = 50$.

3.3. Vergleich verschiedener Fächer

Zum Abschluss dieses Kapitels über den Einfluss verschiedener Fehler auf die resultierenden h-Indizes soll in diesem Abschnitt beleuchtet werden, wie sich die Ergebnisse ändern,

Mathematik		Computerwissenschaften	
ID	*Subject Category*	ID	*Subject Category*: Computer Science,
13	Mathematics	16	Artificial Intelligence
109	Mathematics*	33	Software Engineering
135	Statistics & Probability	39	Hardware & Architecture
227	Mathematics, Applied	55	Computer Science*
336	Mathematics, Interdisciplinary	211	Information Systems
	Applications	225	Interdisciplinary Applications
		255	Cybernetics
		274	Theory & Methods
Ingenieurwissenschaften		**Onkologie**	
ID	*Subject Category*: Engineering,	ID	*Subject Category*:
5	Engineering*	4	Oncology*
		100	Oncology

Tabelle 3.2.: Die in diesem Abschnitt verwendeten Definitionen der verschiedenen untersuchten Fächer. Die mit einem Stern gekennzeichneten Kategorien sind keine traditionellen *Subject Categories* aus dem *Web of Science*, sondern erweiterte Kategorien aus der Datenbank des Kompetenzzentrums Bibliometrie.

wenn statt der Physik andere Fachdisziplinen betrachtet werden. Vorgestellt werden dabei Ergebnisse für deutsche Hochschulpublikationen der Mathematik, ausgewählter Ingenieurwissenschaften, der Onkologie und der Computerwissenschaften. Tabelle 3.2 gibt dabei einen Überblick über die zur Definition der Fächer verwendeten *Subject Categories* des *Web of Science*.

In Tabelle 3.3 sind zu allen untersuchten Fächern einige wichtige Kennzahlen zusammengefasst. Diese Zahlen zeigen, wie die einzelnen Fächer sich aus bibliometrischer Sicht unterscheiden und helfen damit, die Ergebnisse dieses Abschnittes zu interpretieren und zu verstehen. Wichtig ist vor allem die durchschnittliche Anzahl von Zitationen pro Publikation. Diese unterscheiden sich zum Teil erheblich und bewegen sich von nur gut 10%

	Physik	Mathematik	Ingenieur-wissenschaften	Onkologie	Computer-wissenschaften
Gesamtzahl der Publikationen	254832	77167	91608	44015	47244
Gesamtzahl der Zitationen	6094029	788548	1353802	1328846	589516
Zitationen pro Publikation	23,91	10,22	14,78	30,19	12,48
Anteil unzitierter Publikationen	8,5%	6,3%	6,3%	1,6%	4,0%

Tabelle 3.3.: Einige wichtige Kennzahlen der verschiedenen Fächer.

in der Mathematik bis zu über 30% in der Onkologie. Auch bei dem Anteil unzitierter Publikationen sind diese beiden Fächer die positiven und negativen Extreme. Begründet liegen diese Unterschiede in den verschiedenen Fächerkulturen: in der Mathematik wird relativ wenig publiziert (die Gesamtzahl an Publikationen in der Mathematik ist weniger als ein Drittel der Publikationszahlen in der Physik) und dabei auch wenig zitiert, die Medizin gilt hingegen als sehr publikationsstark mit vielen Referenzen pro Publikation.[14] [15]

Die gefundenen Ergebnisse für die relativen gemittelten Fehler der h-Indizes für fehlende Publikationen und fehlende Zitationen sind in den beiden Abbildungen 3.10 und 3.11 in Abhängigkeit der Fehlerwahrscheinlichkeit für verschiedenen Werte von M und für die Anzahl von Konfigurationen $M_M = N = 500$, über die gemittelt wird, dargestellt.

[14] Zitt, Ramanana-Rahary und Bassecoulard, „Relativity of citation performance and excellence measures: From cross-field to cross-scale effects of field-normalisation“, S. 374.

[15] Adam, „Citation analysis: The counting house“, S. 727.

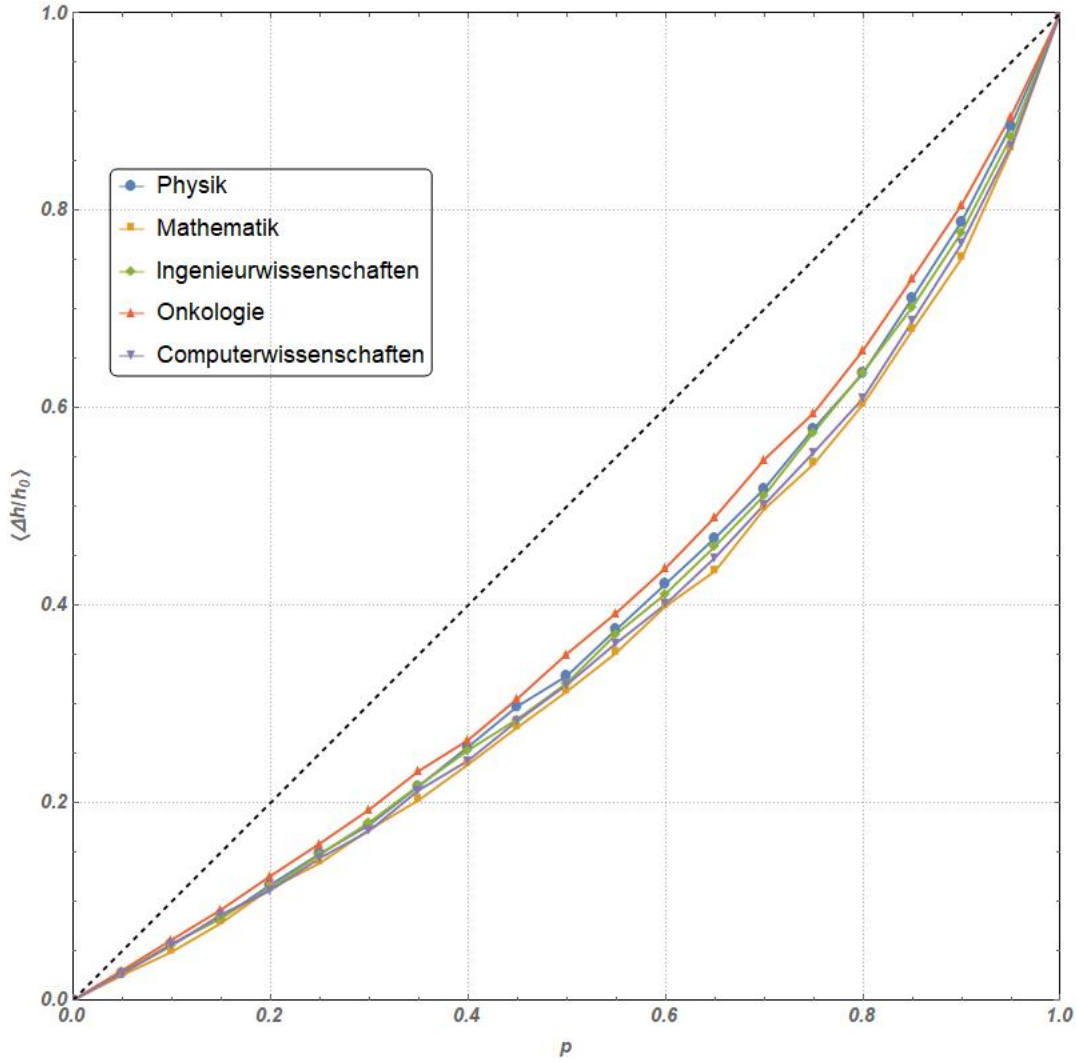

Abbildung 3.10.: Abhängigkeit des relativen gemittelten Fehlers des h-Index von der Fachdisziplin und der Fehlerwahrscheinlichkeit p für den Fall fehlenden Publikationen.

Hier ergibt sich ein ähnliches Verhalten der Kurven wie in den vorherigen Abschnitten: während der Fehler für festes p in Abbildung 3.10 in der Onkologie am größten und in der Mathematik am kleinsten ist (die genaue Reihenfolge ist aufsteigend: Mathematik, Computerwissenschaften, Ingenieurwissenschaften, Physik und Onkologie), ergibt sich in Abbildung 3.11 das umgekehrte Verhalten. Die Verläufe der Kurven zu den übrigen Fächern lassen sich hingegen kaum signifikant unterscheiden. Für die Mathematik ist der Unterschied zwischen den beiden Fehlerarten damit am geringsten, für die Onkologie am größten. Interessant ist die Beobachtung, dass hier die Reihenfolge mit den

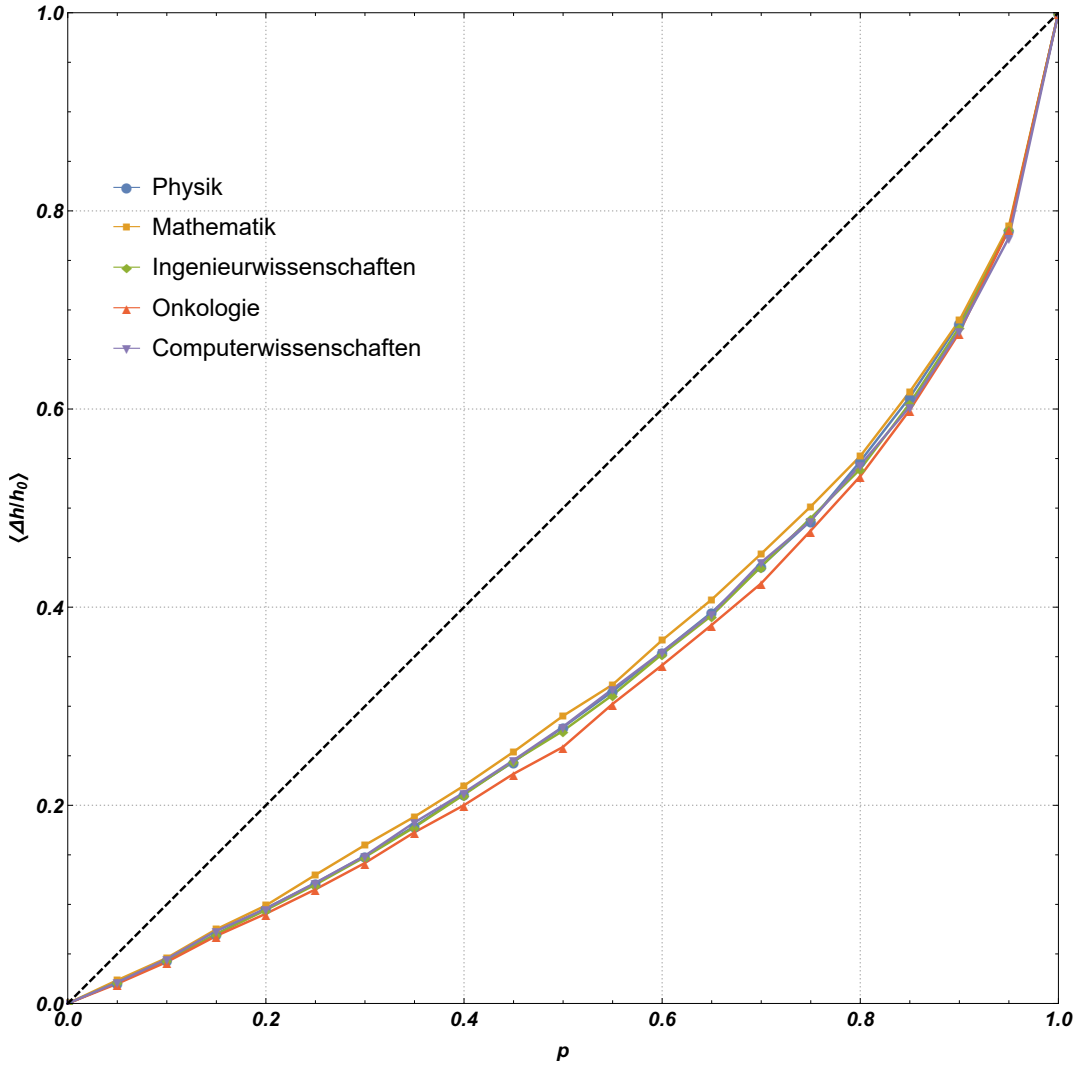

Abbildung 3.11.: Abhängigkeit des relativen gemittelten Fehlers des h-Index von der Fachdisziplin und der Fehlerwahrscheinlichkeit p für den Fall fehlenden Zitationen.

durchschnittlichen Zitationszahlen aus Tabelle 3.3 übereinstimmt.

Insgesamt sind die Unterschiede zwischen den Fächern jedoch eher gering und werden in der Realität deutlich davon überlagert, dass die Fächer eine stark unterschiedliche Verteilung der h-Indizes bei festem M aufweisen. Dies ist in Abbildung 3.12 exemplarisch für die Physik und die Mathematik für $M = 25$ und 100 sowie $M_M = 100000$ Mittelungen dargestellt: in der Physik liegen die Maxima der Verteilungen bei 11 und 25, in der Mathematik bei 9 und 16.

Für alle in diesem Kapitel gefundenen Ergebnisse zum h-Index ist zu beachten, dass

sie einen deutlichen Bias enthalten, da nur Fehler betrachtet werden, die den h-Index

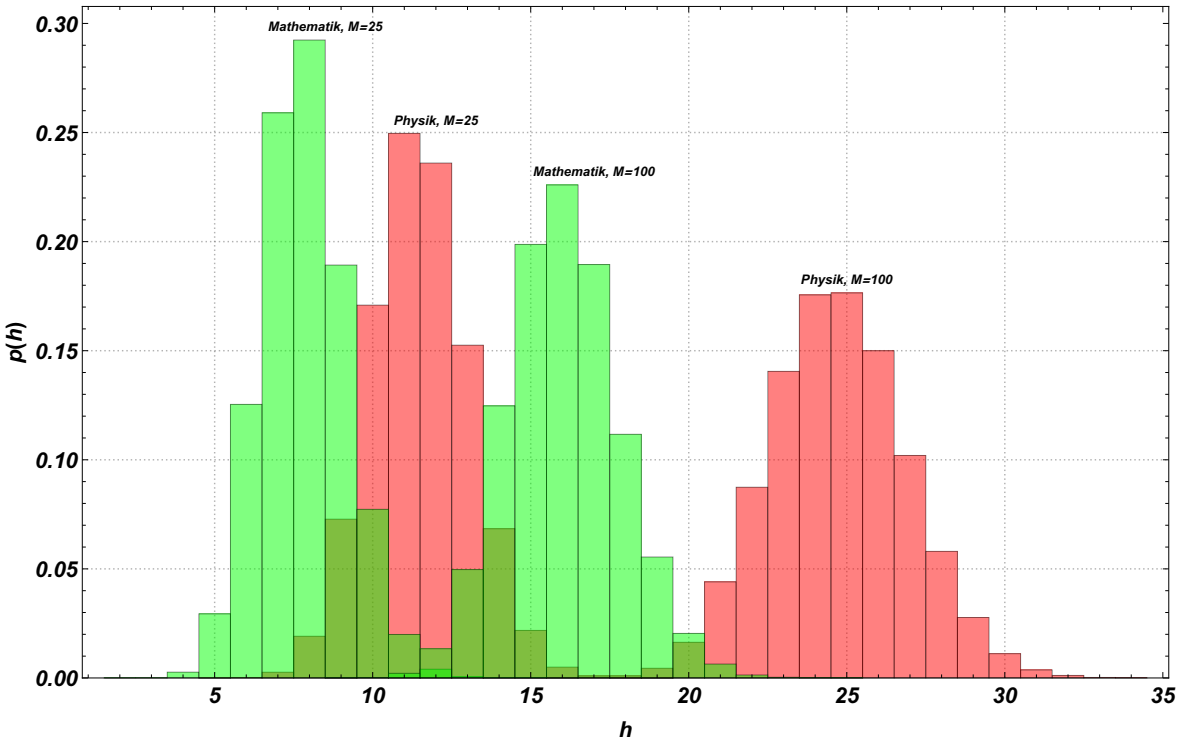

Abbildung 3.12.: Vergleich der Verteilungen der h-Indizes in der Physik und der Mathematik für $M = 25$ und $M = 100$ bei $M_M = 100000$ Mittelungen.

ausschließlich verkleinern können, nicht aber vergrößern. In realistischen Systemen, beziehungsweise bei der Berechnung von h-Indizes tatsächlich existierender Personen, werden die Fehler *im Mittel* also eher kleiner sein. Jedoch ist dabei zu beachten, dass dies aufgrund der Schiefe der Fehlerverteilungen wirklich nur im Mittel gilt. In Einzelfällen kann die Realität stark davon abweichen.

4. Normalisierte Metriken

Seit ihrer Einführung stehen einfache bibliometrische Indikatoren ohne Normalisierung wie der h-Index und *Journal Impact Factor* in der Kritik. Besonders wird dabei auf die häufig fehlende Vergleichbarkeit der Kennzahlen zwischen verschiedenen Fächern, oft sogar innerhalb einer Disziplin, hingewiesen.[1]

Daher wurden bald verbesserte Indikatoren entwickelt, welche dieses Problem umgehen sollten. Dies sollte dadurch erreicht werden, dass bei der Berechnung der Kennzahlen nicht nur die Publikationen der betrachteten Einheit berücksichtigt werden, sondern auch die einer Vergleichsgruppe, welche als Referenz herangezogen wird. Auf diese Weise sollen bibliometrische Kennzahlen vergleichbar werden. Für diese Normalisierung gibt es verschiedene Möglichkeiten und inzwischen existiert eine Reihe von Arbeiten zu normalisierten Indikatoren. In diesem Abschnitt dieser Arbeit sollen zwei normalisierte Indikatoren auf ihre Abhängigkeit von Fehlern in der Datengrundlage untersucht werden, die unterschiedlicher Verfahren zur Normalisierung verwenden: Dies ist zum einen der J-Faktor, der 2009 von Ball, Mittermaier und Tunger vorgeschlagen wurde[2]. Hier werden die Zeitschriften, in der die betrachteten Publikationen erschienen sind, zum Vergleich herangezogen. Zum anderen wird im zweiten Teil dieses Kapitels ein feldnormalisierter Indikator untersucht, der auf dem von van Raan vorgeschlagenen *Crown Indicator*[3] beruht und die fachliche Zuordnung der Publikationen im *Web of Science*, die *Subject*

[1] Glänzel, „On the Opportunities and Limitations of the H-index“.

[2] Ball, Mittermaier und Tunger, „Creation of journal-based publication profiles of scientific institutions — A methodology for the interdisciplinary comparison of scientific research based on the J-factor“.

[3] van Raan, „Measuring Science“, S. 30.

Categories, zur Normalisierung benutzt. Der J-Faktor hat gegenüber dem feldnormalisierten Indikator den Vorteil, dass jede Publikation nur einer Zeitschrift zugeordnet werden kann und somit eindeutig definiert ist, mit welchen Publikationen diese verglichen werden muss. Mehrfachzählungen einzelner Publikationen aufgrund einer Zuordnung zu mehreren Klassen treten hier nicht auf.
Wie auch der h-Index können diese normalisierten Indikatoren für beliebige Publikationssätze bestimmt werden. Dies können einzelne Personen sein, aber auch Arbeitsgruppen oder ganze Institutionen und Länder. Die Untersuchungen dieser Arbeit sollen exemplarisch am konkreten Beispiel der Universität Duisburg-Essen durchgeführt werden.

4.1. Normalisierung auf Basis der Zeitschriften: Der J-Faktor

Der J-Faktor wird immer relativ zu einer Vergleichsgruppe, dem sogenannten Standard, berechnet. Für jede Zeitschrift, in der an der untersuchten Institution im betrachteten Zeitraum publiziert wurde, wird das Verhältnis aus den Zitationsraten der Institution und der Vergleichsgruppe gebildet und mit einem Wichtungsfaktor – der Anzahl der Publikationen der Institution in der jeweiligen Zeitschrift, normiert auf die Gesamtzahl der untersuchten Publikationen – multipliziert. Summiert man anschließend über alle Zeitschriften, so erhält man eine relative Zitationsrate J, welche die relative Wahrnehmung der Publikationen der Institution in Vergleich zum gewählten Standard angibt. Dabei werden Publikations- und Zitationsgewohnheiten auch für interdisziplinäre Forschung berücksichtigt und man erhält einen Indikator, der fachübergreifend vergleichbar ist.[4]

[4] Ball, Mittermaier und Tunger, „Creation of journal-based publication profiles of scientific institutions — A methodology for the interdisciplinary comparison of scientific research based on the J-factor“, S. 1.

Die Formel für den J-Faktor lautet dann

$$J(I,R) = \sum_S \frac{\mathrm{CPP}_I(S)}{\mathrm{CPP}_R(S)} \cdot \frac{p_I(S)}{p_{I,\mathrm{ges}}}, \tag{4.1}$$

wobei

- $\mathrm{CPP}_I(S)$ die durchschnittliche Zitationsrate der Publikationen der Institution in der Zeitschrift S,
- $\mathrm{CPP}_R(S)$ das Analogon für die Publikationen des gewählte Standards,
- $p_I(S)$ die Anzahl der Publikationen der Institution in der Zeitschrift S und
- $p_{I,\mathrm{ges}}$ die Gesamtzahl der Publikationen der Institution bezeichnen.

Selbstzitationen werden dabei nicht ausgeschlossen, was verschiedene Gründe hat: zum einen ist es nicht klar, wie diese eindeutig (automatisiert) identifiziert werden sollen, zum anderen wurde inzwischen gezeigt, dass diese die Ergebnisse nicht verfälschen[5]. Zudem ist es sinnvoll, nicht nur zwischen Publikationen verschiedener Zeitschriften zu differenzieren, sondern auch zwischen verschiedenen Dokumenttypen zu unterscheiden. Wie im vorherigen Abschnitt 3 bei der Analyse des Einflusses von Fehlern in der Datengrundlage auf den h-Index, werden hier wieder nur Publikationen des Typs „Article“ und „Review“ betrachtet. Daher muss die Summe über alle Zeitschriften in Gleichung 4.1 ersetzt werden durch eine Summe über alle Zeitschriften und über alle Dokumenttypen. Entsprechend sind werden die durchschnittlichen Zitationsraten und die Publikationszahlen abhängig von der Zeitschrift und dem Dokumenttyp (DT). Wie auch bei der Wahl der betrachteten Dokumenttypen hat man bei der Wahl des Standards eine gewisse Freiheit. Will man zum Beispiel den gesamten Publikationsoutput eines Landes bewerten, so ist es sinnvoll, alle in den Zeitschriften publizierten Arbeiten als Standard zu wählen. Betrachtet man hingegen nur eine einzelne Institution, so kann es auch sinnvoll sein, als Standard alle Publikationen aus dem selben Land oder aller Institutionen desselben Typs zu betrachten. In den meisten Fällen werden die Publikationen der untersuchten

[5] Glänzel, „Seven Myths in Bibliometrics About facts and fiction in quantitative science studies“.

Einheit Teilmenge des Standards sein, was für die Fehleranalyse von besonderer Bedeutung ist, da so Fehler an mehreren Stellen in die Berechnung des J-Faktors eingehen können.

Eine weitere Besonderheit tritt immer dann auf, wenn keine der Publikationen aus dem Standard, also in einer Zeitschrift und eines Dokumenttyps im betrachteten Zeitraum, zitiert wurde. Dann gilt $\mathrm{CPP}_R(S, DT) = \mathrm{CPP}_I(S, DT) = 0$. In Gleichung 4.1 führt dies zu einem Summanden „0/0“ der nicht definiert ist. Es ist jedoch sinnvoll, diese Summanden auf eins zu setzen, da die betrachtete Institution und der gewählte Standard die gleiche Zitationsrate aufweisen und somit keine von beiden Gruppen bevorzugt wird.

Für den J-Faktor soll im Folgenden exemplarisch am Beispiel der Publikationen der Universität Duisburg-Essen aus dem 10-Jahres-Zeitraum von 2007 bis 2016 der Einfluss von statistischen Fehlern in der Datengrundlage untersucht werden. Die Vorgehensweise ist dabei wieder analog zu der in Abschnitt 3. Dabei soll sich jedoch auf die Art von Fehlern beschränkt werden, bei der einzelne Publikationen fehlen. Für jede Publikation wird dabei eine Wahrscheinlichkeit $0 \leq p \leq 1$ angenommen, mit der diese in der Datengrundlage fehlt. Dabei ist zu beachten, dass dies nicht nur für die Publikationen der untersuchten Institution notwendig ist, sondern auch für die Publikationen des als Referenz gewählten Standards. Insbesondere müssen Publikationen, welche für die Institution gelöscht werden, auch im zugehörigen Standard entfernt werden. Fehlt also eine Publikation der Institution in der Zeitschrift S_i, so geht dieser Fehler an mehreren Stellen in Gleichung 4.1 ein, da sich hier dann alle Faktoren $\mathrm{CPP}_I(S_i)$, $\mathrm{CPP}_R(S_i)$, $p_I(S_i)$ und $p_{I,\mathrm{ges}}$ des zugehörigen Summanden in der Summe über S ändern. Das macht die Implementierung gegenüber der Analyse des h-Index deutlich aufwändiger.[6] Zudem erschwert der große Umfang der Daten, welche berücksichtigt werden müssen, die Analyse: Tabelle 4.1 gibt einige Kennzahlen der in diesem Abschnitt benötigten Datengrundlagen sowie den aus Gleichung 4.1 bestimmten, ungestörten J-Faktor an. Interessant ist dabei die

[6] Trotzdem kann der Effekt, der dadurch entsteht, dass wenn eine Publikation aus der Datenbasis entfernt wird, eigentlich auch die zugehörigen Zitationszahlen der übrigen Publikationen modifiziert werden müssten, wieder nicht berücksichtigt werden. Dies würde den Aufwand zusätzlich enorm steigern.

Jahr	Publikationen DUE	Zitationen DUE	Mittlere Zitationszahl DUE	Anzahl Zeitschriften	Publikationen Standard	Zitationen Standard	Mittlere Zitationszahl Standard	J-Faktor
2007	1447	40725	28,14	794	250287	8491173	33,93	1,11905
2008	1462	44735	30,60	831	268360	8855561	33,00	1,0451
2009	1659	48540	29,26	896	285011	8462631	29,69	1,06763
2010	1685	49582	29,43	969	302289	8114797	26,84	1,14791
2011	1711	39152	22,88	966	331729	7474970	22,53	1,14136
2012	1765	35390	20,05	989	356654	6569184	18,42	1,16628
2013	1938	30506	15,74	1036	381459	5226948	13,7	1,23189
2014	1858	21700	11,68	1053	407861	3979029	9,76	1,17965
2015	2052	17338	8,45	1117	436238	2435259	5,581	1,22275
2016	2109	3897	1,85	1127	449197	681060	1,52	1,25521

Tabelle 4.1.: Einige Kennzahlen der in diesem Abschnitt verwendeten Datengrundlagen, aufgeschlüsselt nach Jahren: Die ersten drei Spalten zeigen die Anzahl der Publikationen, die Anzahl der Zitationen sowie die mittlere Zitationszahl pro Publikation, jeweils für die Universität Duisburg-Essen. Die Zahl der Publikationen pro Jahr nimmt kontinuierlich zu, die der Zitierungen, insbesondere für die späteren Jahre, deutlich ab. Letzteres ist dadurch zu erklären, dass die meisten Zitationen erst einige Jahre nach Erscheinen der Arbeit anfallen. Die nächste Spalte zeigt die Anzahl der Zeitschriften, in denen in diesem Jahr publiziert wurde. Auche diese Zahl nimmt über die Jahre deutlich zu. Danach folgen die Publikations- und Zitationszahlen des als Referenz gewählten Standards. In diesem Fall sind das alle Artikel desselben Jahres aus allen Zeitschriften, in denen in diesem Jahr Publiziert wurde. Die letzte Spalte gibt den aus den Daten nach Formel 4.1 bestimmten J-Faktor an (siehe Haupttext).

Beobachtung, dass der J-Faktor für die Universität Duisburg-Essen für alle Jahre größer als eins ist, obwohl ihre durchschnittliche Zitationsrate unter der des Standards liegt. Dies ist jedoch kein Widerspruch, da bei der Berechnung des J-Faktors Publikationen aus verschiedenen Zeitschriften unterschiedlich gewichtet werden.

Während es bei der Berechnung des ungestörten J-Faktors ausreicht, die Zitationsraten

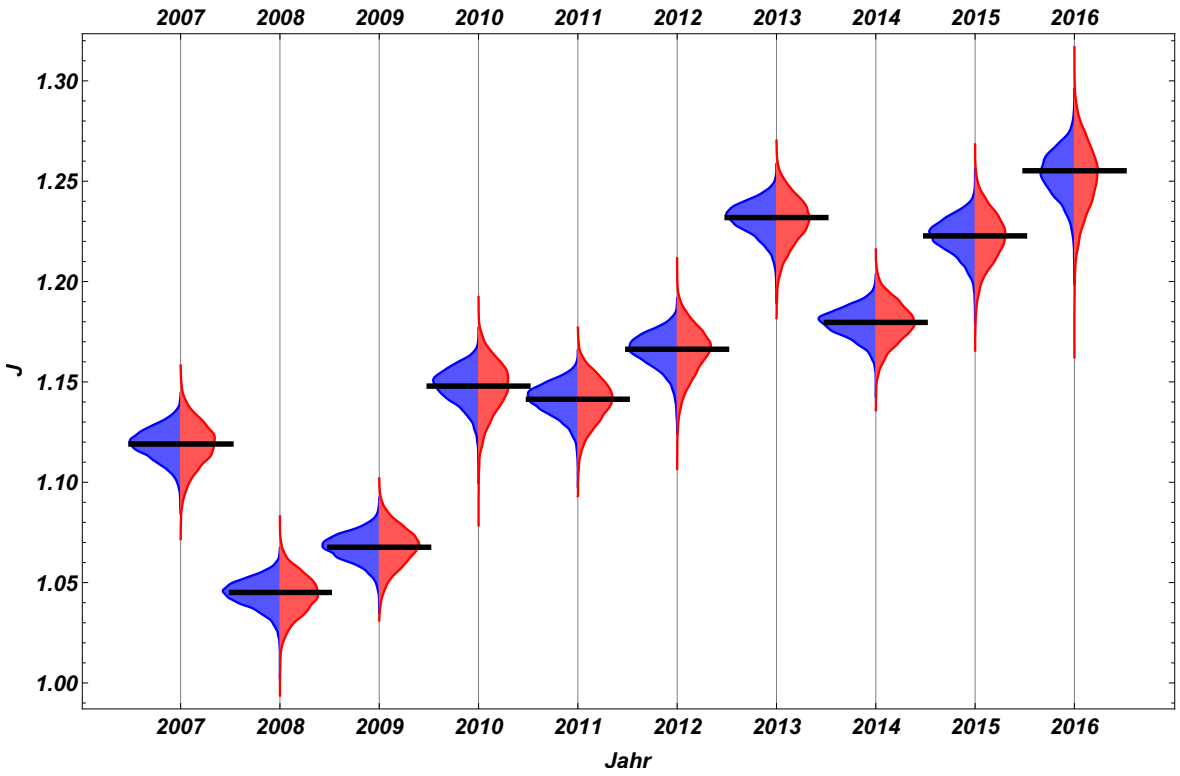

Abbildung 4.1.: Verteilungen der fehlerbehafteten J-Faktoren für die Fehlerwahrscheinlichkeiten $p = 0{,}05$ (blau) und $p = 0{,}1$ (rot) für die untersuchten Jahre in einem Violinplot. Die schwarzen Linien Zeigen den ungestörten J-Faktor des entsprechenden Jahres.

und Publikationszahlen für die benötigten Zeitschriften abzurufen, müssen für die Fehleranalyse alle Publikationen in den benötigten Zeitschriften – sowohl die der Universität Duisburg-Essen, als auch die des betrachten Standards – abgerufen werden. Nur so ist es möglich, einzelne Publikationen aus der Datenbasis zu löschen. Benötigt werden dabei das Publikationsjahr, Identifikatoren der Zeitschrift und des Artikels, die Zitationszahl sowie der Dokumenttyp. Nach Tabelle 4.1 macht das knapp 3,5 Millionen Datensätze, die für die Analyse benötigt werden, wodurch der Rechenaufwand gegenüber der Ana-

lyse des h-Index im vorherigen Kapitel stark ansteigt.
In Anhang A.2 sind Auszüge aus dem verwendeten Mathematica-Quellcode angegeben, mit dem die über SQL-Abfragen aus der Datenbank des Kompetenzzentrums Bibliometrie gewonnenen Daten innerhalb von Mathematica strukturiert werden können und daraus der J-Faktor – sowohl für eine endliche Fehlerwahrscheinlichkeit in der Datenbasis, als auch für das ungestörte System bestimmt werden kann.
Die mit diesem Programmcode erhaltenen Ergebnisse sind in den Abbildungen 4.1-4.4 graphisch dargestellt. Berechnet wird dafür der J-Faktor der Universität Duisburg-Essen für feste Werte von Fehlerwahrscheinlichkeit p und Jahr, jeweils für 10000 Fehlerkonfigurationen, bei denen jede Publikation aus dem Datensatz der Universität Duisburg-Essen und dem Standard zufällig mit der Wahrscheinlichkeit p gelöscht wurde.
Die so erhaltenen Verteilungen sind für zwei verschiedene Werte $p = 0{,}05$ und $p = 0{,}1$ abhängig vom Jahr in Abbildung 4.1 als Violinplot dargestellt. Für jedes Jahr sind somit die beiden Wahrscheinlichkeitsverteilungen der fehlerbehafteten J-Faktoren für zwei Fehlerwahrscheinlichkeiten gegenübergestellt. Man erkennt in dieser Darstellung deutlich die Aufweitung der Fehlerbalken für größer werdende p. Zudem gibt diese Darstellung ein Gefühl dafür, inwieweit J-Faktoren aus verschiedenen Jahren, welche sich teilweise nur leicht unterscheiden – wie zum Beispiel die aus den Jahren 2010 und 2011 – , signifikant voneinander unterschieden werden können. In dem Beispiel der beiden Jahre 2010 und 2011 kann keine verlässliche Aussage darüber getroffen werden, welcher J-Faktor tatsächlich größer ist, da der Überlapp der beiden Verteilungen sehr groß ist.
Aus den jeweiligen Zufallsverteilungen kann die Wahrscheinlichkeit berechnet werden, dass zwei aus fehlerhaften Datenbasen bestimmte Werte von dem J-Faktor mit $J_1 < J_2$ in Wirklichkeit – das heißt unter Annahme einer exakten Datenbasis – die Eigenschaft $J_1 > J_2$ erfüllen. Diese Wahrscheinlichkeit ist natürlich für aus bibliometrischen Indikatoren abgeleitete Positionen in Rankings von großem Interesse.[7]
Die vorausgegangene Aussage soll im Folgenden an einem Beispiel quantifiziert werden:

[7] Raan, „Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods“.

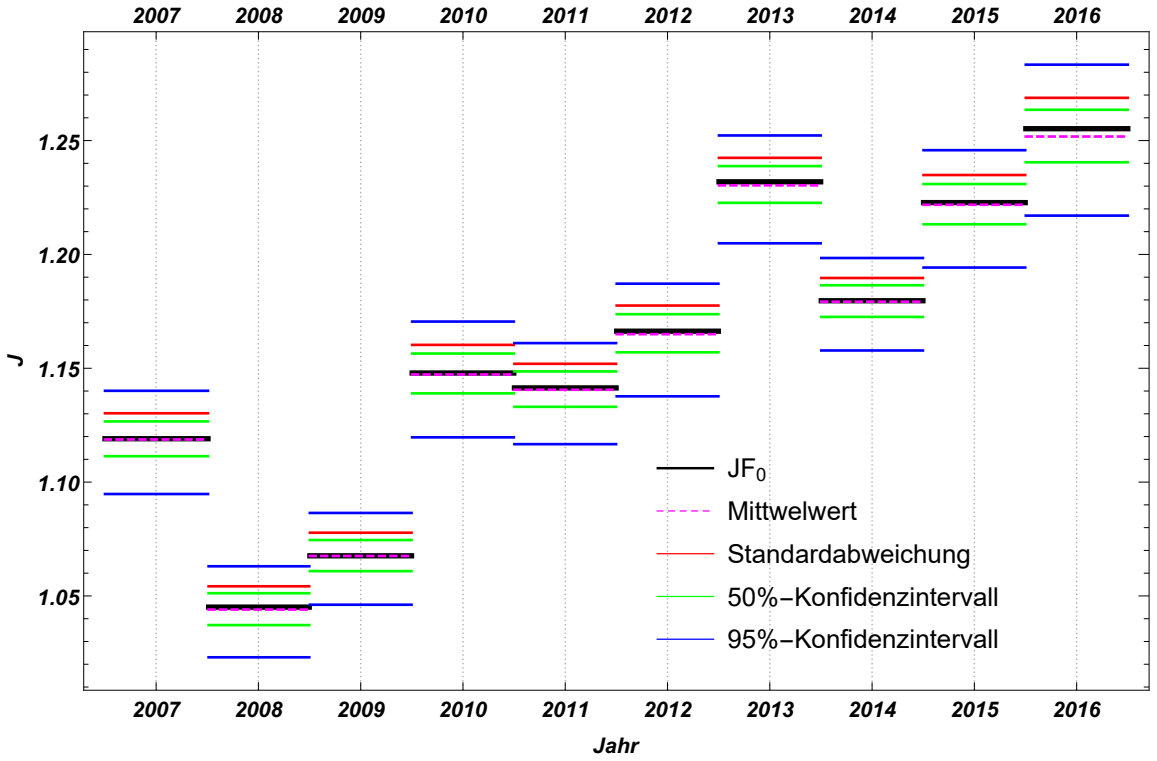

Abbildung 4.2.: Abschätzung der Zuverlässigkeit der J-Faktoren für verschiedene Jahre: Die schwarzen Linien zeigen den exakten Wert des J-Faktors, die strichlierten Linien den Erwartungswert der Verteilung des J-Faktors mit $p = 0{,}1$. Die roten Linienpaare geben die Breite der Standardabweichung um den Erwartungswert herum an, die grünen und blauen Paare die 50%- beziehungsweise 95%-Konfidenzintervalle (siehe Haupttext).

Die Wahrscheinlichkeit, dass für zwei zufällig bestimmte Werte der beiden J-Faktoren der Jahre 2010 und 2011 die Eigenschaft $J_{\text{err}}(2010) < J_{\text{err}}(2011)$ gilt (was für die ungestörten F-Faktoren gemäß Tabelle 4.1 eigentlich nicht zutrifft), lässt sich aus den beiden zugehörigen Wahrscheinlichkeitsverteilungen bei $p = 0{,}1$ – $w_{2010}(x)$ und $w_{2010}(x)$, welche in Abbildung 4.1 dargestellt sind – über ein Faltungsintegral berechnen. Es gilt

$$W(J_{\text{err}}(2010) < J_{\text{err}}(2011)) = \int_{-\infty}^{0} \mathrm{d}z \int_{0}^{\infty} \mathrm{d}x\, w_{2010}(x) w_{2011}(x - z) = 0{,}346, \qquad (4.2)$$

Für die Wahrscheinlichkeit $W(J_{\text{err}}(2010) < J_{\text{err}}(2011))$ ergibt sich also durch numerisches Ausführen der Integration ein Wert 34,6%, das heißt mit dieser Wahrscheinlichkeit bekommt man aus je einer zufälligen Fehlerkonfigurationen zwei J-Faktoren der

	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
2007	–		0,1%	5,4%	9,1%	0,5%				
2008		–	5,4%							
2009	0,1%	5,4%	–							
2010	5,4%			–	34,6%	16,4%		2,5%		
2011	9,1%			34,6%	–	7,9%		0,6%		
2012	0,5%			16,4%	7,9%	–		19,3%	0,1%	
2013							–	0,1%	31,7%	15,4%
2014				2,5%	0,6%	19,3%	0,1%	–	0,7%	
2015						0,1%	31,7%	0,7%	–	8,4%
2016							15,4%		8,4%	–

Tabelle 4.2.: Wahrscheinlichkeiten für Positionswechel der beiden J-Faktoren für alle Paare von Jahren in Prozent für $p = 0{,}1$.

beiden Jahre heraus, die eine andere Reihenfolge in der Sortierung nach ihrer Größe aufweisen als die Maxima der Verteilungen es nahelegen würden. Die übrigen nichtverschwindenden Wahrscheinlichkeiten, dass bei zwei J-Faktoren verschiedener Jahre die beiden Werte für J ihre Position im Ranking tauschen ist in Tabelle 4.2 angegeben. Qualitativ deutlich wird dies auch in Abbildung 4.2, in der die Fehlerintervalle für verschiedene Jahre dargestellt sind. Während die Mittelwerte der Verteilungen jeweils gut mit den exakten Werten J_0 für den J-Faktor übereinstimmen, zeigen die Positionen der Linien der Standardabweichungen und der 50%- und 95%-Konfidenzintervalle eine deutliche Abweichung der fehlerbehafteten J-Faktoren von dem exakten Wert. Das bedeutet, dass ein aus einer fehlerhaften Datenbasis bestimmter Wert für den J-Faktor mit einer Wahrscheinlichkeit von 50% innerhalb der rot gekennzeichneten Intervalle liegt und mit einer Wahrscheinlichkeit von 95% innerhalb der blauen. Die Werte für die Jahre 2010 und 2011 sind also tatsächlich kaum zu unterscheiden. Auch zwischen den anderen Jahren gibt es einen deutlichen Überlapp. Es ist jedoch zu beachten, dass dies nur für die angenommene Fehlerwahrscheinlichkeit von $p = 0{,}1$ gilt. Ist der tatsächliche Feh-

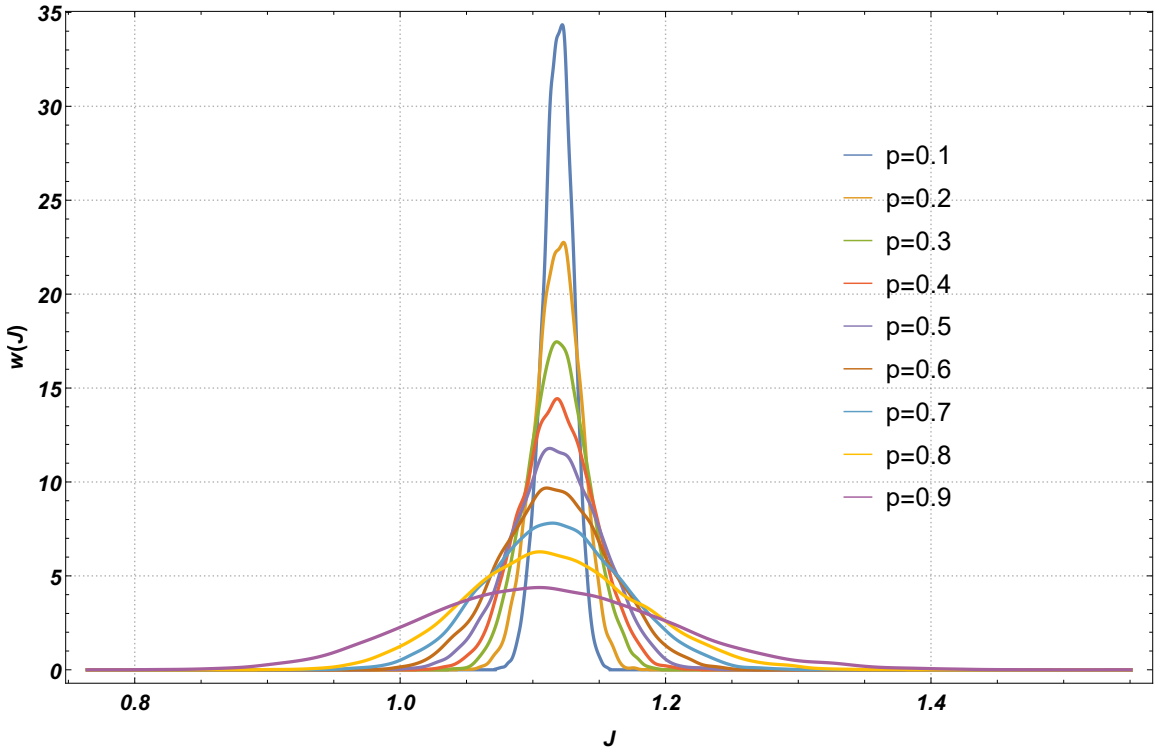

Abbildung 4.3.: Verbreiterung der Wahrscheinlichkeitsverteilungen des J-Faktors für das Jahr 2007 mit steigender Fehlerwahrscheinlichkeit p.

ler kleiner, so wird auch der Überlapp kleiner und die Unterscheidbarkeit größer. Ist die Fehlerwahrscheinlichkeit jedoch größer, so verstärkt sich der Effekt noch deutlich und die Wahrscheinlichkeiten in Tabelle 4.2 nehmen stark zu. Insbesondere treten dann auch für andere Kombinationen zweier Jahre nichtverschwindende Tauschwahrscheinlichkeiten auf. Die Verbreiterung der Wahrscheinlichkeitsverteilungen lässt sich in den beiden Abbildungen 4.3 und4.4 gut erkennen: Abbildung 4.3 zeigt die Wahrscheinlichkeitsverteilungen der fehlerbehafteten J-Faktoren exemplarisch für das Jahr 2007 und verschiedene Fehlerwahrscheinlichkeiten p. Für kleine Werte von p sind die Kurven noch schmal, weiten sich mit wachsendem p aber immer mehr auf. Gleichzeitig verschiebt sich der Mittelwert leicht zu kleineren Werten von J. Deutlicher wird dies in Abbildung 4.4, wo die Lage des Mittelwerts, der Standardabweichungen sowie der 50%- und 95%-Konfidenzintervalle aufgetragen sind. Zum Vergleich ist zusätzlich der exakte Wert J_0 als schwarze Linie eingezeichnet. Alle Intervalle verbreitern sich deutlich mit wach-

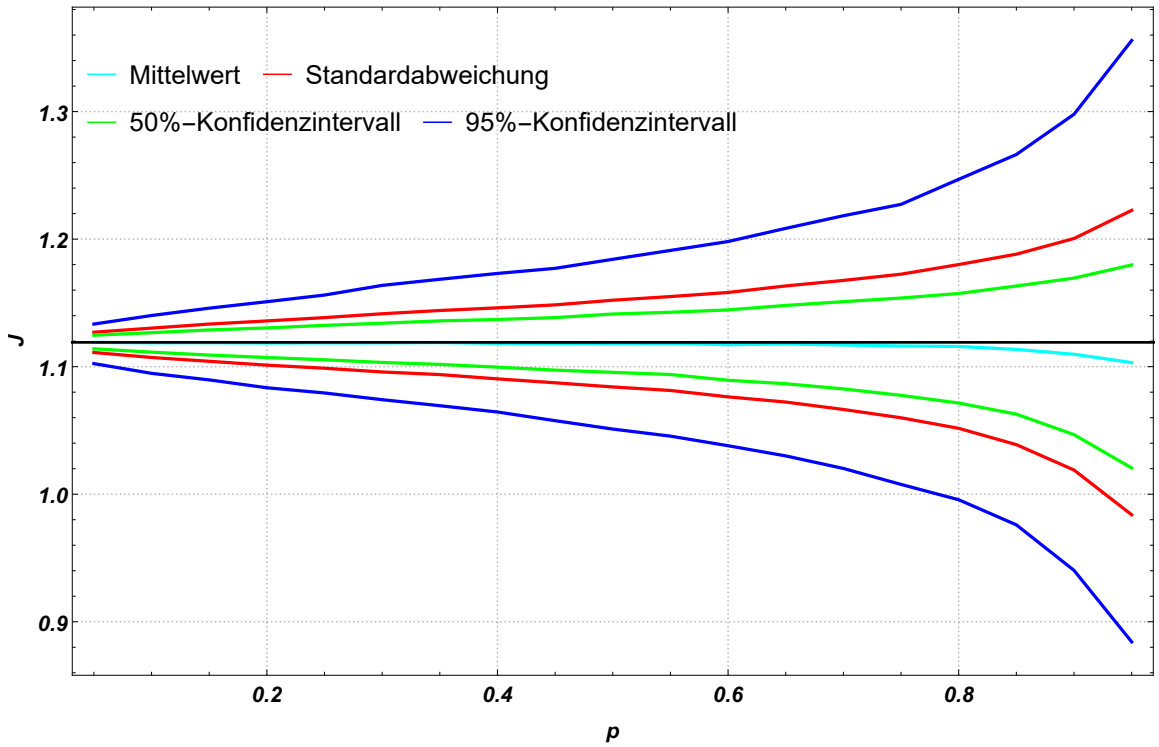

Abbildung 4.4.: Lage des Mittelwerts, der Standardabweichungen sowie der 50%- und 95%-Konfidenzintervalle der Wahrscheinlichkeitsverteilungen des J-Faktors für das Jahr 2007 abhängig von der Fehlerwahrscheinlichkeit p.

sendem p. Für Werte von $p \lesssim 0{,}6$ verhalten sich die Kurven noch annähernd linear, danach wachsen sie überproportional.

4.2. Normalisierung auf Basis der Fächer: Der Crown Indicator

Wie bereits im letzten Abschnitt erwähnt ist die Normalisierung auf Basis der Zeitschriften, in denen die untersuchten Publikationen erschienen sind, nicht die die einzige Möglichkeit, wenn man einen Publikationssatz gegen einen Standard vergleichen will. Deutlich häufiger wird dabei die fachliche Zuordnung der einzelnen Publikationen verwendet. Diese fachliche Kategorisierung erfolgt meist in der Zitationsdatenbank. Im *Web*

of Science sind das die sogenannten *Subject Categories*, von denen jeder verzeichneten Publikation bis zu vier zugeordnet sind.[8]

In diesem Abschnitt soll eine solcher Indikator mit Feldnormalisierung untersucht werden. Als Beispiel wird ein Indikator behandelt, der an den *Crown Indicator* von van Raan angelehnt ist.[9] Dabei gibt es jedoch ein paar Unterschiede, die vor allem technisch begründet sind. Um diese Unterschiede nachvollziehen zu können, soll im Folgenden kurz erläutert werden, wie nach van Raan der *Crown Indicator* berechnet wird: Der *Crown Indicator* ist das Verhältnis zweier bibliometrischer Kennzahlen, der Zitationsrate CPP der untersuchten Einheit und dem *mean Field Citation Score* FCSm,

$$\frac{\mathrm{CPP}}{\mathrm{FCSm}}. \tag{4.3}$$

CPP ergibt sich einfach daraus, dass die Gesamtzahl aller Zitationen, die die Publikationen der untersuchten Institution im betrachteten Zeitraum erhalten haben, durch die Anzahl der Publikationen geteilt wird. FCSm hingegen wird berechnet, indem für jede Publikation der Institution die Zitationsrate aller im selben Zeitraum und im selben Fach publizierten Artikel bestimmt wird und diese Zitationsraten über alle Publikationen gemittelt werden. Dies lässt sich durch die Gleichung

$$\frac{\mathrm{CPP}}{\mathrm{FCSm}} = \frac{\sum_{i=1}^{n} c_i/n}{\sum_{i=1}^{n} e_i/n} \tag{4.4}$$

ausdrücken.[10] Dabei ist c_i die Anzahl der Zitationen der i-ten Publikation der Institution und e_i die Zitationsrate aller Publikationen mit derselben fachlichen Zuordnung wie die i-te Publikation. Summiert wird über alle n untersuchten Publikationen. Nach van Raan werden dabei jedoch nicht alle Zitationen gezählt, sondern nur die in einem 5-Jahres-Fenster nach der Veröffentlichung des Artikels. Zudem sollen Selbstzitationen dabei ausgeschlossen werden. Beides würde die Analyse in dieser Arbeit deutlich verkomplizieren, zum einen, weil zusätzlich zu den Zitationszahlen die Erscheinungsjahre der zitierenden Publikationen abgerufen werden müssten, zum anderen, weil nicht klar

[8] Eine Übersicht darüber findet sich unter Clarivate Analytics, *Web of Science Core Collection Help*.

[9] van Raan, „Measuring Science“, S. 30.

[10] Gleichung (1) in Waltman, Eck u. a., „Towards a new crown indicator: an empirical analysis“, S. 469.

ist, wie Selbstzitationen zuverlässig ausgeschlossen werden können. Zudem würde die Vergleichbarkeit zu den Ergebnissen zum J-Faktor nicht mehr gegeben sein, da dort diese Einschränkungen nicht vorgenommen wurden. Daher soll auch in diesem Abschnitt darauf verzichtet werden.

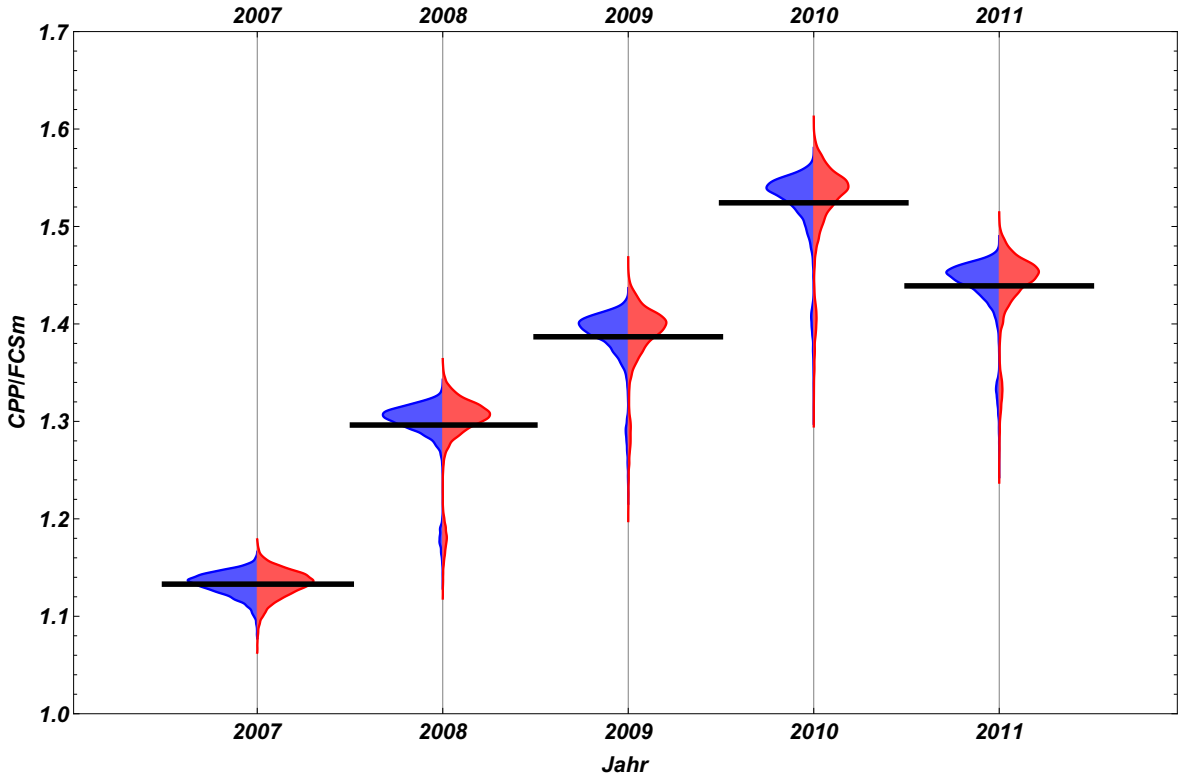

Abbildung 4.5.: Verteilungen der fehlerbehafteten Werte von CPP/FCSm für die Fehlerwahrscheinlichkeiten $p = 0{,}05$ (blau) und $p = 0{,}1$ (rot) für die untersuchten Jahre in einem Violinplot. Die schwarzen Linien Zeigen den ungestörten Wert von CPP/FCSm des entsprechenden Jahres.

Die eigentliche Analyse und Auswertung zum *Crown Indicator* erfolgt nun völlig analog zu der aus dem letzten Abschnitt beim J-Faktor. Insbesondere werden die selben Größen analysiert und die selben Abhängigkeiten untersucht. Daher werden die Ausführungen hier kurzgehalten und vor allem auf die Unterschiede zum J-Faktor eingegangen.
Ein erster wesentlicher Unterschied ist in der Menge der Daten begründet, die hier ausgewertet werden müssen: will man wieder die Jahre 2007–2016 analysieren, so müssen

insgesamt gut 28000 Datensätze für die Universität Duisburg-Essen prozessiert werden. Für jedes dieser Jahre enthalten diese Publikationen Vertreter aus knapp 200 verschiedenen *Subject Categories*. Insgesamt sind das 236 verschiedene *Subject Categories* in dem 10-Jahres-Zeitraum. Zur Berechnung des *Crown Indicators* benötigt man dann alle Publikationen in diesen Kategorien für alle betrachteten Jahre. Zusammen wären das über 20 Millionen Datensätze, was die Numerik sehr zeitaufwändig macht und die Anzahl der Fehlerkonfigurationen, welche gerechnet werden können, stark einschränkt. Aus diesem Grund sollen die Untersuchungen in diesem Abschnitt auf die Jahre 2007–2011 eingeschränkt werden. Zudem werden Zuordnungen zu mehreren *Subject Categories* derselben Publikation vernachlässigt und nur die Kategorie mit dem niedrigsten Identifikator verwendet werden. Dies vereinfacht die Implementierung erheblich.

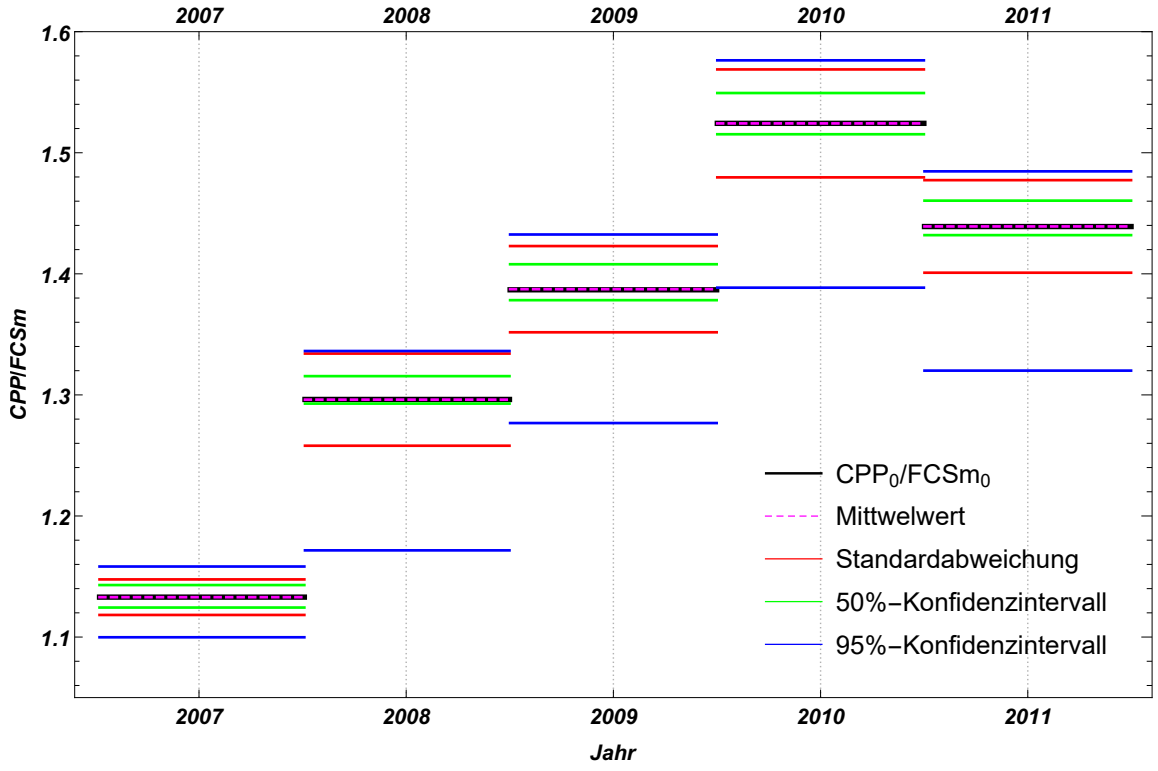

Abbildung 4.6.: Abschätzung der Zuverlässigkeit von $\mathrm{CPP/FCSm}$ für verschiedene Jahre: Die schwarzen Linien zeigen den exakten Wert von $\mathrm{CPP/FCSm}$, die strichlierten Linien den Erwartungswert der Verteilung von $\mathrm{CPP/FCSm}$ mit $p = 0{,}1$. Die roten Linienpaare geben die Breite der Standardabweichung um den Erwartungswert herum an, die grünen und blauen Paare die 50%- beziehungsweise 95%-Konfidenzintervalle (siehe Haupttext).

	2007	2008	2009	2010	2011
2007	–	0,2			
2008	0,2	–	5,3		1,5
2009		5,3	–	3,4	10,1
2010			3,4	–	7,4
2011		1,5	10,1	7,4	–

Tabelle 4.3.: Wahrscheinlichkeiten für Positionswechel der beiden Werte von CPP/FCSm für alle Paare von Jahren in Prozent für $p = 0{,}1$.

Abbildung 4.6 enthält wieder die Wahrscheinlichkeitsverteilungen für CPP/FCSm für $p = 0{,}05$ und 0,1 und verschiedene Jahre als Violinplots, ermittelt aus 10000 Fehlerkonfigurationen. Auffällig für die Jahre 2008-2011 ist dabei eine Lage der Maxima der Verteilungen deutlich über dem jeweiligen exakten Wert für CPP/FCSm. Hervorgerufen wird dies durch ein Nebenmaximum in den Verteilungen, welches zwar klein aber deutlich ausgeprägt unterhalb des exakten Wertes liegt. Dies ist ein bemerkenswerter Effekt, der eine nähere Untersuchung erfordert. Ob es sich dabei um ein Artefakt der Simulationen und einer zu niedrigen Anzahl von Fehlerkonfigurationen handelt, ist vorerst unklar. Interessant wäre es auch zu analysieren, welche Fehlerkonfigurationen zu Werten von CPP/FCSm in den Nebenmaxima führen. Dies könnte für die Interpretation des Phänomens hilfreich sein.

Auswirkungen haben diese Nebenmaxima auch auf die Größe der 95%-Konfidenzintervalle, wie man deutlich in Abbildung 4.6 erkennen kann. Dies führt dazu, dass es zu einem deutlichen Überlapp der Intervalle kommt, obwohl die ungestörten Werte CPP/FCSm für die verschiedenen Jahre deutlich unterschieden werden können. Tabelle 4.3 gibt wieder die Wahrscheinlichkeiten an, mit der zwei Jahre ihre Reihenfolge im Ranking nach CPP/FCSm tauschen würden. Die Werte sind hier deutlich kleiner als beim J-Faktor, was aber vor allem an dem größeren Abstand zwischen den Werten der einzelnen Jahre liegt. Aufgrund der Unsicherheit, ob die Untersuchung von 10000 Fehlerkonfigurationen hier für eine gute Statistik ausreichend sind, müssen diese Zahlen

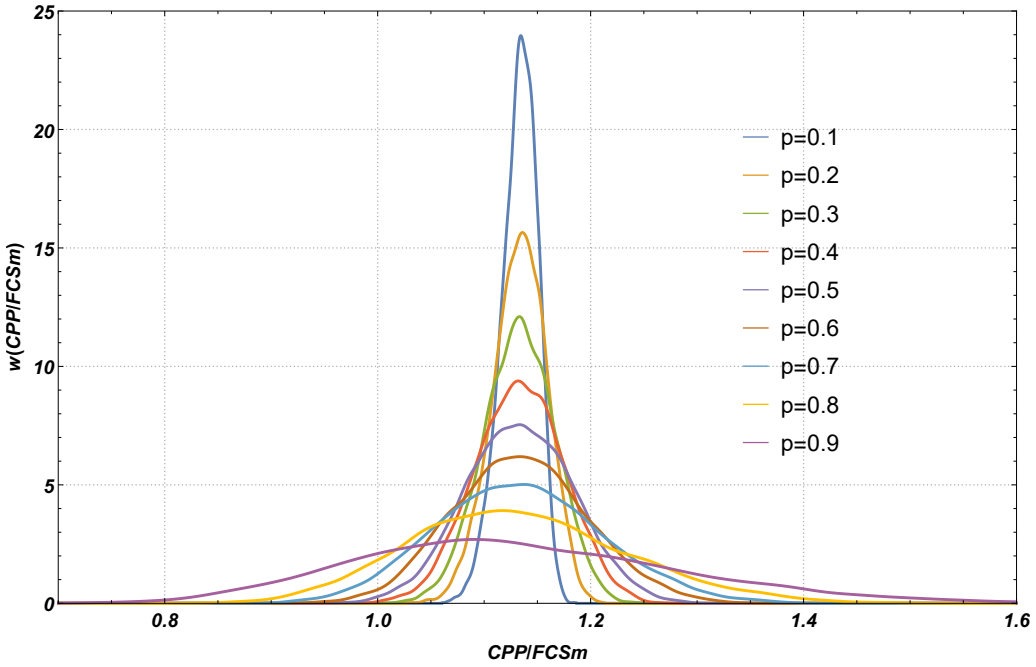

Abbildung 4.7.: Verbreiterung der Wahrscheinlichkeitsverteilungen von $\mathrm{CPP/FCSm}$ für das Jahr 2007 mit steigender Fehlerwahrscheinlichkeit p.

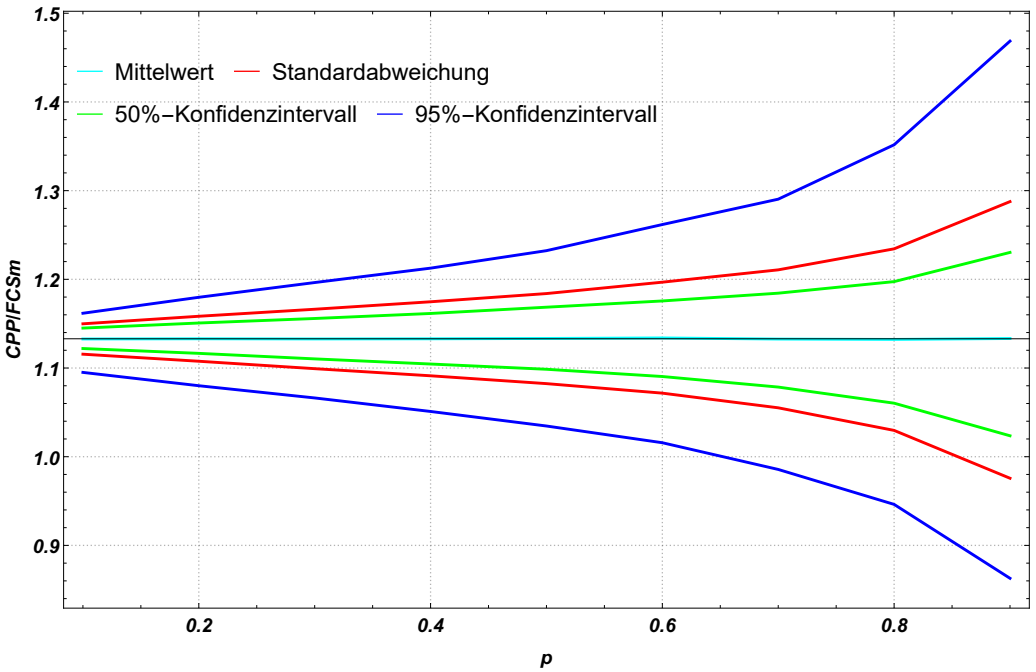

Abbildung 4.8.: Lage des Mittelwerts, der Standardabweichungen sowie der 50%- und 95%-Konfidenzintervalle der Wahrscheinlichkeitsverteilungen von $\mathrm{CPP/FCSm}$ für das Jahr 2007 abhängig von der Fehlerwahrscheinlichkeit p.

kritisch hinterfragt werden. Einen Hinweis darauf, dass eigentlich noch mehr Konfigurationen hätten analysiert werden müssen, gibt auch Abbildung 4.7. Hier erkennt man deutlich eine Struktur innerhalb des Maximums der Verteilungen (zum Beispiel bei $p = 0{,}4$), was darauf hindeutet, dass die Ergebnisse noch nicht ausreichen statistisch konvergiert sind.

Zum Abschluss dieses Abschnitts zeigt Abbildung 4.8 die Abhängigkeiten der verschiedenen statistischen Intervalle in Abhängigkeit von der Fehlerwahrscheinlichkeit. Die Verläufe zeigen hier qualitativ das gleiche Verhalten wie beim J-Faktor.

5. Diskussion und Ausblick

5.1. Diskussion

Im Rahmen dieser Masterarbeit wurde eine Methode entwickelt, mit der die Auswirkungen von Fehlern in der Datengrundlage auf verschiedene bibliometrische Indikatoren untersucht werden können. Dabei werden in eine als exakt angenommene Datenbasis, welche ungestörte Werte als Referenz liefert, statistisch Fehler verschiedener Art eingestreut und berechnet, wie sich die aus der ungestörten und der fehlerhaften Datenbasis bestimmten Werte unterscheiden. Implementiert werden können dabei Fehler, bei der einzelne Publikationen oder einzelne Zitationen mit einer vorgegebenen Wahrscheinlichkeit gelöscht werden beziehungsweise fehlen. Dabei müssen die gestörten bibliometrischen Indikatoren für ausreichend viele Fehlerkonfigurationen bestimmt werden, damit die Ergebnisse statistisch konvergieren und die Wahrscheinlichkeitsverteilungen der gestörten Kennzahlen ausgewertet werden können. Aus diesen Verteilungen können dann über die üblichen statistischen Größen wie Standardabweichung und Konfidenzintervalle Fehlerbalken bestimmt werden. Zudem lassen sich daraus Wahrscheinlichkeiten berechnen, mit denen zwei fehlerbehaftete Kennzahlen die Positionen in Rankings tauschen.
Angewendet wurde die Methode zunächst auf den h-Index, wobei hier keine realen Personenprofile, sondern künstlich generierte Publikationssets fiktiver Wissenschaftler verwendet wurden.
Anschließend wurde die Methode auf zwei normalisierte Indikatoren – den J-Faktor und

den *Crown Indicator* – angewendet. Dabei wurden keine künstlich erzeugten Publikationssets verwendet, sondern das konkrete Beispiel der Universität Duisburg-Essen.
Während die untersuchten Fehler bei h-Index immer dazu führen, dass der Erwartungswert des gestörten Wertes kleiner als der des ungestörten ist, verhalten sich die normalisierten Metriken da grundlegend anders: dadurch, dass die Fehler sowohl in den Publikationen der untersuchten Institution als auch in der als Referenz verwendeten Datenbasis implementiert werden, streuen hier die fehlerbehafteten Indikatoren in beide Richtungen um den ungestörten Wert. Der Mittelwert stimmt dabei recht genau mit dem exakten Wert überein, die endliche Breite der Wahrscheinlichkeitsverteilungen führen dann aber dazu, dass trotzdem nur bestimmte Wahrscheinlichkeiten dafür angegeben werden können, dass der „richtige“ Indikator in einem gewissen Intervall um den berechneten gestörten Wert liegt.

Erwähnenswert dabei ist auch, dass Fehler auch durch unklare Definitionen der Indikatoren entstehen können. Diese führen dazu, dass Werte aus verschiedenen Arbeiten nicht miteinander verglichen werden können, weil aus den Publikationen nicht hervorgeht, wie die Indikatoren in allen Details berechnet werden sollen. Unklar dabei ist häufig der Umgang mit Selbstzitationen, welche Dokumententypen berücksichtigt werden sollten und die genaue Datenbasis. Dies führt zu eigentlich vermeidbaren Fehlern und erschwert die Vergleichbarkeit zwischen Arbeiten verschiedener Autoren.

5.2. Ausblick

Es wäre wünschenswert, wenn die in dieser Arbeit vorgestellten Analysen einen Anstoß dazu geben würden, dass sich vermehrt mit dem Thema Fehler in bibliometrischen Analysen beschäftigt wird. Unmittelbar aus dem Kontext dieser Arbeit ergeben sich bereits einige weitere Fragestellungen, die den Rahmen einer Masterarbeit gesprengt hätten, aber lohnenswert für weitere Untersuchungen wären:

In dem Abschnitt zur Untersuchung des h-Index wurden nur künstliche Autorenprofile untersucht, die dadurch gewonnen wurden, dass zufällig M Publikationen aus einer Grundgesamtheit gewählt wurden. Diese Methode könnte man noch deutlich verfeinern, wenn man nicht nur die Publikationszahl M, sondern auch die durchschnittliche Anzahl von Publikationen pro Jahr und die Dauer der Aktiven Zeit sowie wie lange diese zurückliegt, berücksichtigen würde. Auf diese Weise könnten deutlich realistischere – wenn auch weiterhin künstlich generierte – Autorenprofile erzeugt werden. Wünschenswert wäre natürlich auch eine Erweiterung auf reale Autorenprofile.
Zudem wäre es sinnvoll, auch Fehler zu implementieren die dazu führen würden, dass der h-Index ansteigen kann. In der vorliegenden Arbeit wurden hingegen nur Fehler betrachtet, bei denen Publikationen oder Zitationen fehlten.
Interessant wäre es auch, wie sich die Wahrscheinlichkeitsverteilungen für den fehlerbehafteten h-Index bei festem ungestörtem h_0 (statt festem M) verhalten. Durch die Analyse der Standardabweichungen und Konfidenzintervalle könne man wieder analysieren, mit welcher Wahrscheinlichkeit zwei h-Indizes mit der Relation $h_{\text{err}}^{(1)} < h_{\text{err}}^{(2)}$ für die ungestörten („echten“) h-Indizes $h_0^{(1)} > h_0^{(2)}$ gilt und die daraus folgenden Personenrankings verfälscht würden.

Die Untersuchungen zum J-Faktor wurden hier nur an dem Beispiel der Universität Duisburg-Essen durchgeführt. Eine Erweiterung auf andere Institutionen wäre durchaus wünschenswert. Insbesondere die Analyse der Abhängigkeit der Fehler von der Institutionsgröße könnte eine wichtige Fragestellung sein, da dies allgemeinere Abschätzungen von Fehlerbalken des J-Faktors ermöglichen könnte.

Aufgrund der Größe der Universität Duisburg-Essen, der entsprechend hohen Publikationszahlen und insbesondere aufgrund der großen Anzahl von zu berücksichtigenden *Subject Categories* wurden bei der Untersuchung der feldnormalisierten Indikatoren Mehrfachzuordnungen zu verschiedene *Subject Categories* ignoriert. Dies sollte den Aufwand der numerischen Rechnungen, die im Rahmen dieser Arbeit durchgeführt wurden, begrenzen. Es wäre also sinnvoll, eine kleinere Institution zur Untersuchung zu wählen,

und anhand dieser die Unterschiede zu analysieren, die durch eine Berücksichtigung der Mehrfachzuordnungen entstehen. Publikationen mit mehreren Klassifikationen gehen dabei häufiger in den Indikator ein, dadurch wäre zu erwarten, dass der Fehler beim Entfernen dieser Publikationen verstärkt wird. Das Verhalten der Wahrscheinlichkeitsverteilungen von CPP/FCSm, ein Nebenmaximum unterhalb des Mittelwertes auszubilden, bedarf sicherlich weiter Untersuchungen. Auch hierfür würden sich kleinere System anbieten.

Weitere Fragestellungen wären die ausführlichere Untersuchung des Einflusses von Fehlern in der Datenbasis auf Positionen in Rankings sowie die Erweiterung der Analysen auf Metriken, die aus der Netzwerktheorie stammen.
Es gibt also noch eine ganze Reihe von offenen Fragestellungen, die es wert wären, näher erläutert zu werden.

— Anhang —

A. Mathematica-Quellcode

Die meisten Simulationen und Berechnungen wurden mit Mathematica durchgeführt. Dies hat den Vorteil, dass innerhalb einer Umgebung sowohl SQL-Abfragen an die Datenbank des Kompetenzzentrums geschickt werden können, als auch auf umfangreiche vorimplementierte mathematische Routinen für aufwändige Rechnungen sowie zur graphischen Darstellung der Ergebnisse zurückgegriffen werden kann.
Die wesentlichen Teil des Quellcodes der verwendeten Programme sollen in diesem Abschnitt kurz vorgestellt werden.

A.1. Anbindung an die Oracle-Datenbank des Kompetenzzenrrums

Mathematica 10 hat keinen Treiber für Oracle-Datenbanken vorinstalliert. Dies kann jedoch mit ein paar Schritten nachgeholt werden:
Zunächst muss das Paket „DatabaseLink“ geladen werden und ein neues Verzeichnis für die JAR-Datei des JDBC-Treibers erstellt werden. Den passenden Treiber bekommt man von Oracle. Dieser wird in das neu angelegte Verzeichnis abgelegt. Anschließend muss noch eine Datei mit der JDBC Treiberkonfiguration erstellt werden, damit der Treiber von Mathematica erkannt wird.
Nach diesen Schritten kann eine Verbindung zur Datenbank erstellt werden und mit dem Befehl SQLExecute SQL-Abfragen ausgeführt werden:

```
Needs["DatabaseLink`"]

$jarDirectory = "C:\\Users\\...\\Mathematica\\Applications\\Oracle\\Java"

C:\Users\...\Mathematica\Applications\Oracle\Java

SystemOpen[$jarDirectory]

$configDirectory =
 CreateDirectory@
  FileNameJoin@{$UserBaseDirectory, "Applications", "Oracle", "DatabaseResources"}


Export[FileNameJoin@{$configDirectory, "Oracle.m"},
 JDBCDriver["Name" → "Oracle", "Driver" → "oracle.jdbc.driver.OracleDriver",
  "Protocol" → "jdbc:oracle:thin:@", "Version" → 1], "Text"]

$connection = OpenSQLConnection[JDBC["Oracle", "[Servername der Datenbank]"],
  "Username" → "...", "Password" → "..."]

SQLExecute[$connection,
 "SELECT * FROM WOS_B_2017.KB_INST WHERE WOS_B_2017.KB_INST.NAME LIKE '%Duisburg%'"]
```

A.2. Analyse des h-Index

Die Variable physikdat enthält eine Tabelle mit den Spalten PK_ITEMS, PUBYEAR, COUNT. Die Funktion hindphysik berechnet den h-Index für eine zufällige Auswahl von n Publikationen aus physikdat:

```
maxphysik = Length[datphysik]

hindphysik[n_] :=
 Select[Reverse[Sort[RandomChoice[datphysik, n][[All, 3]]] - Range[1, n], # ≥ 0 &] // Length
```

Die Funktionen pubconfig und errconfig erzeugen zufällige Publikations- und Fehlerkonfigurationen:

```
Clear[pubconfig]
pubconfig[M_, MM_] := Table[RandomSample[Table[i, {i, 1, max}], M], MM]

Clear[errconfig]
errconfig[M_, p_, NN_] :=
 Table[RandomSample[Table[i, {i, 1, M}],
   RandomVariate[HypergeometricDistribution[M, Floor[p max], max]]], NN]
```

Die Funktion hindexerrphysik1a berechnet die h-Indizes für eine feste Anzahl fehlender Publikationen in der Grundgesamtheit physikdat:

```
hindexerrphysik1a[p_, M_, NN_, MM_] := Block[{pconf, econf, sample, sampleerr, h0, herr},
   pconf = pubconfig[M, MM]; econf = errconfig[M, p, NN];
   sample = Table[Part[datphysik[[All, 3]], pconf[[m]]], {m, 1, MM}];
   h0 = Table[Count[Sort[Part[datphysik[[All, 3]], pconf[[m]]], Greater] - Range[M],
      u_ /; u ≥ 0], {m, 1, MM}];
   herr = Table[Count[Sort[ReplacePart[sample[[m]], {econf[[n]}]^T → 0], Greater] - Range[M],
      u_ /; u ≥ 0], {m, 1, MM}, {n, 1, NN}];
   {h0, herr}];
```

Die Funktion hindexerrphysik1b berechnet die h-Indizes für eine feste Fehlerwahrscheinlichkeit für jede Publikationen:

```
hindexerrphysik1b[p_, M_, NN_] := Block[{samp, conf, h0, herr},
  samp = RandomSample[datphysik[[All, 3]], M];
  herr = {};
  h0 = Count[Sort[samp, Greater] - Range[M], u_ /; u ≥ 0];
  Do[conf = Table[If[Random[] - p > 0, 1, 0], M];
   herr = Append[herr, Count[Sort[samp conf, Greater] - Range[M], u_ /; u ≥ 0]];, NN];
  {h0, herr}
 ]
```

Die Funktion hindexerrphysik2b berechnet die h-Indizes für eine feste Fehlerwahrscheinlichkeit für jede Zitation:

```
hindexerrphysik2b[p_, M_, NN_] := Block[{samp, conf, h0, herr},
  samp = RandomSample[datphysik[[All, 3]], M];
  herr = {};
  h0 = Count[Sort[samp, Greater] - Range[M], u_ /; u ≥ 0];
  Do[conf = Table[If[Random[] - p > 0, 1, 0], M];
   herr = Append[herr, Count[Sort[samp conf, Greater] - Range[M], u_ /; u ≥ 0]];, NN];
  {h0, herr}
 ]
```

A.3. Analyse des J-Faktors

Die Variable pubude enthält 5 Spalten mit den Einträgen PK_ITEMS, COUNT, PUBYEAR, FK_SOURCES, DOCTYPE; jdata2007 fünf Spalten mit den Einträgen PUBYEAR, FK_SOURCES, DOCTYPE, PK_ITEMS, COUNT. Die Daten werden aufbereitet, mit der Funktion JFerr[p,year] kann dann für gegebenes Jahr und Fehlerwahrscheinlichkeit der fehlerbehaftete J-Faktor berechnet werden.

```
journals[year_] := Select[pubude, (#[[3]] == year) &][[All, 4]] // DeleteDuplicates
Do[
 (jdata[jid, doctype, 2007] = Select[jdata2007, ( #[[3]] == doctype && #[[2]] == jid) &][[
    All, {4, 5}]]), {doctype, {"Article", "Review"}}, {jid, Take[journals[2007], All]}]
Do[
 Do[
  (instdata[jid, doctype, year] =
    Select[pubude, ( #[[5]] == doctype && #[[4]] == jid && #[[3]] == year) &][[All, {1, 2}]]),
  {doctype, {"Article", "Review"}}, {jid, journals[year]}], {year, 2007, 2016}]
Block[{list}, Do[
  Do[list = Tally[Flatten[{jdata[jid, doctype, year], instdata[jid, doctype, year]}, 1]];
   joineddata[jid, doctype, year] = {list[[All, 1, 2]], list[[All, 2]]}^T,
   {doctype, {"Article", "Review"}}, {jid, journals[year]}], {year, 2007, 2016}]]
Do[totaldue[year] = Select[pubude // DeleteDuplicates, ( #[[3]] == year) &] // Length,
 {year, 2007, 2016}]


 JFerr[p_, year_] := Block[{inst, journal, J, randata, rand, total}, J = 0; total = 0;
   Do[rand = Table[If[Random[] - p > 0, 1, 0], Length[joineddata[jid, doctype, year]]];
    randata[jid, doctype] = Select[{joineddata[jid, doctype, year], rand}^T, #[[2]] == 1 &][[
      All, 1]];
    total = total + Count[randata[jid, doctype][[All, 2]], 2]; , {jid, journals[year]},
    {doctype, {"Article", "Review"}}];
   Do[inst = Select[randata[jid, doctype], #[[2]] == 2 &][[All, 1]];
    rand = rand + Length[inst];
    journal = randata[jid, doctype][[All, 1]];
    J =
     J + If[Length[inst] == 0, 0, If[ Total[journal] == 0, 1,
           Total[inst] / Length[inst]
          Total[journal] / Length[journal]]] Length[inst]
                                              total
                                                  // N, {jid, journals[year]},
    {doctype, {"Article", "Review"}}];
   J]
```

A.4. Analyse des Crown Indicators

Die Daten werden in die beiden Variablen rawdata und rawdatadue gespeichert und haben die vier Spalten PK_ITEMS, FK_CLASSIFICATIONS, DOCTYPE, COUNT. scall enthält alle *Subject Categories* des jeweiligen Jahres. In joineddata werden die Daten zusammengeführt und strukturiert. Mit den Funktionen Crown0 und Crownerr werden die ungestörten und gestörten Werte für den *Crown Indicator* berechnet.

```
In[23]:= Do[scall[n] = rawdatadue[n][[All, 2]] // DeleteDuplicates // Sort, {n, Range[2007, 2016]}]
         Flatten[Table[scall[n], {n, Range[2007, 2016]}]] // DeleteDuplicates // Length

In[33]:= Do[Block[{scdata, instdata, list},
           Do[scdata = Select[rawdata[n], (#[[2]] == scid && #[[3]] == doctype) &][[All, {1, 4}]];
            instdata = Select[rawdatadue[n], (#[[2]] == scid && #[[3]] == doctype) &][[All, {1, 4}]];
            list = Tally[Flatten[{scdata, instdata}, 1]];

            (joineddata[n, scid, doctype] = {list[[All, 1, 2]], list[[All, 2]]}^T),
            {scid, scall[n]}, {doctype, {"Article", "Review"}}]], {n, Range[2007, 2016]}]
```

```
In[82]:= Clear[Crown0]
         Crown0[year_] := Crown0[year] = Block[{inst, sc, ci, FCSm},
            ci = 0.;
            FCSm = 0.;
            Do[
             inst = Select[joineddata[year, scid, doctype], #[[2]] == 2 &][[All, 1]];
             sc = joineddata[year, scid, doctype][[All, 1]];
             FCSm = FCSm + If[Length[sc] > 0, Total[sc]/Length[sc] Length[inst], 0];
             ci = ci + Total[inst];

             , {scid, scall[year]}, {doctype, {"Article", "Review"}}];
            ci / FCSm]
```

In[236]:=
```
Clear[Crownerr]
Crownerr[year_, p_] := Block[{inst, sc, ci, rand, randata, FCSm},
  ci = 0.;
  FCSm = 0.;
  Do[rand = Table[If[Random[] - p > 0, 1, 0], Length[joineddata[year, scid, doctype]]];
   randata[scid, doctype] = Select[{joineddata[year, scid, doctype], rand}^T, #[[2]] == 1 &][[
     All, 1]];
   , {scid, scall[year]}, {doctype, {"Article", "Review"}}];
  Do[
   inst = Select[randata[scid, doctype], #[[2]] == 2 &][[All, 1]];
   sc = randata[scid, doctype][[All, 1]];
   FCSm = FCSm + If[Length[sc] > 0, Total[sc]/Length[sc] Length[inst], 0];
   ci = ci + Total[inst];
   , {scid, scall[year]}, {doctype, {"Article", "Review"}}];
  ci / FCSm]
```

Literatur

Adam, David. „Citation analysis: The counting house“. In: *Nature* 415.6873 (Feb. 2002), S. 726–729. DOI: 10.1038/415726a.

Ball, Rafael. „Bibliometrische Dienstleistungen“. In: *Praxishandbuch Bibliotheks?management.* Hrsg. von Rolf Griebel, Hildegard Schäffler und Konstanze Söllner. 6.10. de Gruyter, 2015, S. 556–575. URL: https://epub.uni-regensburg.de/31050/.

Ball, Rafael, Bernhard Mittermaier und Dirk Tunger. „Creation of journal-based publication profiles of scientific institutions — A methodology for the interdisciplinary comparison of scientific research based on the J-factor“. In: *Scientometrics* 81.2 (März 2009), S. 381–392. DOI: 10.1007/s11192-009-2120-5.

Bergstrom, Carl. „Eigenfactor: Measuring the value and prestige of scholarly journals“. In: *College & Research Libraries News* 68.5 (Mai 2007), S. 314–316. DOI: 10.5860/crln.68.5.7804.

Bornmann, Lutz und Loet Leydesdorff. „Skewness of citation impact data and covariates of citation distributions: A large-scale empirical analysis based on Web of Science data“. In: *Journal of Informetrics* 11.1 (2017), S. 164–175. ISSN: 1751-1577. DOI: https://doi.org/10.1016/j.joi.2016.12.001. URL: http://www.sciencedirect.com/science/article/pii/S1751157716303108.

Braun, T. und A. Schubert. „Dimensions of scientometric indicator datafiles“. In: *Scientometrics* 38.1 (Jan. 1997), S. 175–204. DOI: 10.1007/bf02461130.

Brin, S. und L. Page. „The Anatomy of a Large-Scale Hypertextual Web Search Engine“. In: *Seventh International World-Wide Web Conference (WWW 1998)*. 1998. URL: http://ilpubs.stanford.edu:8090/361/.

Clarivate Analytics. *History of Citation Indexing*. URL: https://clarivate.com/essays/history-citation-indexing/ (besucht am 06. 10. 2018).

— *Web of Science Core Collection Help*. URL: https://images.webofknowledge.com/images/help/WOS/hp_subject_category_terms_tasca.html (besucht am 06. 10. 2018).

Courtault, Jean-Michel und Naïla Hayek. „On the Robustness of the h-index: a mathematical approach“. In: *Economics Bulletin* 3.78 (Dez. 2008), S. 1–9. URL: https://halshs.archives-ouvertes.fr/halshs-00446309.

DIN 1319-1 (1995-01-00). *Grundlagen der Meßtechnik - Teil 1: Grundbegriffe*.

Egghe, Leo und Ronald Rousseau. „The Hirsch index of a shifted Lotka function and its relation with the impact factor“. In: *Journal of the American Society for Information Science and Technology* 63.5 (2012), S. 1048–1053. DOI: 10.1002/asi.22617. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.22617. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.22617.

Falagas, Matthew E. u. a. „Comparison of SCImago journal rank indicator with journal impact factor“. In: *The FASEB Journal* 22.8 (Aug. 2008), S. 2623–2628. DOI: 10.1096/fj.08-107938.

Franceschini, Fiorenzo, Domenico Maisano und Luca Mastrogiacomo. „Empirical analysis and classification of database errors in Scopus and Web of Science“. In: *Journal of Informetrics* 10.4 (2016), S. 933–953. ISSN: 1751-1577. DOI: https://doi.org/10.1016/j.joi.2016.07.003. URL: http://www.sciencedirect.com/science/article/pii/S175115771630061X.

Gimpl, Kerstin. „Evaluation von ausgewählten Altmetrics-Diensten für den Einsatz an wissenschaftlichen Bibliotheken“. de. Master Thesis. 2017, S. 78. URL: http://nbn-resolving.de/urn:nbn:de:hbz:79pbc-opus-10341.

Glänzel, Wolfgang. „On the Opportunities and Limitations of the H-index“. In: 1 (Jan. 2006).

Glänzel, Wolfgang. „Seven Myths in Bibliometrics About facts and fiction in quantitative science studies“. In: *COLLNET Journal of Scientometrics and Information Management* 2.1 (2008), S. 9–17. DOI: 10.1080/09737766.2008.10700836. eprint: https://doi.org/10.1080/09737766.2008.10700836. URL: https://doi.org/10.1080/09737766.2008.10700836.

Glänzel, Wolfgang und Koenraad Debackere. „On the opportunities and limitations in using bibliometric indicators in a policy relevant context. Applications, Benefits and Limitations. 2nd Conference of the Central Library“. In: *Bibliometric Analysis in Science and Research*. Hrsg. von R. Ball. Bd. 11. Schriften des Forschungszentrums Jülich : Bibliothek / Library. Forschungszentrum, Zentralbibliothek, 2003, S. 225–236.

Harzing, Anne-Wil und Satu Alakangas. „Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison“. In: *Scientometrics* 106.2 (Feb. 2016), S. 787–804. ISSN: 1588-2861. DOI: 10.1007/s11192-015-1798-9. URL: https://doi.org/10.1007/s11192-015-1798-9.

Haustein, Stefanie und Dirk Tunger. „Sziento- und bibliometrische Verfahren“. In: *Grundlagen der praktischen Information und Dokumentation: Handbuch zur Einführung in die Informationswissenschaft und -praxis*. Hrsg. von Wolfgang Semar Rainer Kuhlen und Dietmar Strauch. DE GRUYTER SAUR, 2013, S. 479–492.

Havemann, Frank. *Einführung in die Bibliometrie*. Humboldt-Universität zu Berlin, Philosophische Fakultät I, 2009. DOI: http://dx.doi.org/10.18452/9432.

Hicks, Diana u. a. „Bibliometrics: The Leiden Manifesto for research metrics“. In: *Nature* 520.7548 (Apr. 2015), S. 429–431. DOI: 10.1038/520429a.

Hirsch, J. E. „An index to quantify an individual's scientific research output“. In: *Proceedings of the National Academy of Sciences* 102.46 (Nov. 2005), S. 16569–16572. DOI: 10.1073/pnas.0507655102.

Hornbostel, Stefan. *Wissenschaftsindikatoren – Bewertungen in der Wissenschaft*. Westdeutscher Verlag, 1997.

Jokić, Maja und Rafael Ball. *Qualität und Quantität wissenschaftlicher Veröffentlichungen : bibliometrische Aspekte der Wissenschaftskommunikation*. ger. Jülich: Forschungszentrum Jülich, 2006. ISBN: 3893364315.

Jovanović, Miloš. „Eine kleine Frühgeschichte der Bibliometrie“. In: *Information - Wissenschaft & Praxis* 63.2 (Jan. 2012). DOI: 10.1515/iwp-2012-0017.

Kompetenzzentrum Bibliometrie. *Dateninfrastruktur*. URL: http://www.forschungsinfo.de/Bibliometrie/index.php?id=infrastruktur (besucht am 06. 10. 2018).

— *Über das Kompetenzzentrum Bibliometrie*. URL: http://www.forschungsinfo.de/bibliometrie/ (besucht am 06. 10. 2018).

Li, Jie u. a. „Citation Analysis: Comparison of Web of Science®, Scopus™, SciFinder®, and Google Scholar“. In: *Journal of Electronic Resources in Medical Libraries* 7.3 (2010), S. 196–217. DOI: 10.1080/15424065.2010.505518. eprint: https://doi.org/10.1080/15424065.2010.505518. URL: https://doi.org/10.1080/15424065.2010.505518.

Mandelbrot, Benoit. „An informational theory of the statistical structure of language“. In: *Communication theory* 84 (1953), S. 486–502.

Minnick, Jennifer. *A closer look at the Journal Impact Factor numerator*. Hrsg. von Clarivate Analytics. 26. Apr. 2017. URL: https://clarivate.com/blog/science-research-connect/closer-look-journal-impact-factor-numerator/ (besucht am 06. 10. 2018).

Moed, H. F. und Th. N. van Leeuwen. „Impact factors can mislead“. In: *Nature* 381.6579 (Mai 1996), S. 186–186. DOI: 10.1038/381186a0.

Olensky, Marlies. „Data accuracy in bibliometric data sources and its impact on citation matching“. Diss. Humboldt-Universität zu Berlin, Philosophische Fakultät I, 2015. DOI: http://dx.doi.org/10.18452/17122.

„Physics and sport“. In: *Physics World* 25.07 (2012), S. 15. URL: http://stacks.iop.org/2058-7058/25/i=07/a=24.

Pritchard, Alan. „Statistical Bibliography or Bibliometrics“. In: *Journal of Documentation* 25.4 (1969), S. 348–349.

Raan, Anthony F. J. van. „Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods“. In: *Scientometrics* 62.1 (Jan. 2005), S. 133–143. DOI: 10.1007/s11192-005-0008-6.

Ramsay, Maurice. „Cycling record“. In: *Physics World* 18.11 (2005), S. 22. URL: http://stacks.iop.org/2058-7058/18/i=11/a=30.

Schreiber, Michael. „A skeptical view on the Hirsch index and its predictive power“. In: *Physica Scripta* 93.10 (Sep. 2018), S. 102501. DOI: 10.1088/1402-4896/aad959.

Silagadze, Z. K. „Citations and the Zipf-Mandelbrot's law“. In: *Complex Syst.* 11 (1997), S. 487–499. arXiv: physics/9901035 [physics].

Thelwall, Mike und Ruth Fairclough. „The accuracy of confidence intervals for field normalised indicators“. In: *Journal of Informetrics* 11.2 (Mai 2017), S. 530–540. DOI: 10.1016/j.joi.2017.03.004.

van Raan, Anthony F. J. „Measuring Science“. In: *Handbook of Quantitative Science and Technology Research*. Springer Netherlands, 2004, S. 19–50. DOI: 10.1007/1-4020-2755-9_2.

Vanclay, Jerome K. „On the robustness of the h-index“. In: *Journal of the American Society for Information Science and Technology* 58.10 (2007), S. 1547–1550. DOI: 10.1002/asi.20616.

Waltman, Ludo und Nees Jan van Eck. „The inconsistency of the h-index“. In: *Journal of the American Society for Information Science and Technology* 63.2 (2012), S. 406–415. DOI: 10.1002/asi.21678. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.21678. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.21678.

Waltman, Ludo, Nees Jan van Eck u. a. „Towards a new crown indicator: an empirical analysis“. In: *Scientometrics* 87.3 (Juni 2011), S. 467–481. ISSN: 1588-2861. DOI: 10.1007/s11192-011-0354-5. URL: https://doi.org/10.1007/s11192-011-0354-5.

Weisstein, Eric W. *Hypergeometric Distribution. From MathWorld–A Wolfram Web Resource.* URL: http://mathworld.wolfram.com/HypergeometricDistribution.html (besucht am 06. 10. 2018).

Zipf, George Kingsley. „The Meaning-Frequency Relationship of Words“. In: *The Journal of General Psychology* 33.2 (Okt. 1945), S. 251–256. DOI: 10.1080/00221309.1945.10544509.

Zitt, Michel, Suzy Ramanana-Rahary und Elise Bassecoulard. „Relativity of citation performance and excellence measures: From cross-field to cross-scale effects of field-normalisation“. In: *Scientometrics* 63.2 (Apr. 2005), S. 373–401. ISSN: 1588-2861. DOI: 10.1007/s11192-005-0218-y. URL: https://doi.org/10.1007/s11192-005-0218-y.

Eigenständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe.
Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Dies gilt auch für Quellen aus eigenen Arbeiten.
Ich versichere, dass ich diese Arbeit oder nicht zitierte Teile daraus vorher nicht in einem anderen Prüfungsverfahren eingereicht habe.
Mir ist bekannt, dass meine Arbeit zum Zwecke eines Plagiatsabgleichs mittels einer Plagiatserkennungssoftware auf ungekennzeichnete Übernahme von fremdem geistigem Eigentum überprüft werden kann.

Duisburg, 09. Oktober 2018 Felix Schmidt